Author:
Zhang, Nina

Title:
A Study of Interprovincial Migration in China Using Extended Gravity Models

A Study of Interprovincial Migration in China Using Extended Gravity Models

Xingna Nina Zhang

A thesis submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Social Science and Law.

School of Geographical Sciences

March 2019

Word count: 71,696

This thesis is dedicated to my parents and supervisors!

Abstract

Regional inequality has been put forward in the literature as a major driver of interprovincial migration in China. However, three major levels of regional inequality – rural and urban, province, and region have not been systematically investigated regarding their effects on interprovincial migration. This thesis chooses the gravity model of migration to investigate this. Moreover, three knowledge gaps persist across many studies using gravity models, including the neglect of different migration types, flow data dependencies and possible non-linear distance decay.

These knowledge gaps call into question the applicability of the standard gravity model. Particularly, the issue of flow reciprocity has been largely ignored by the literature; a significant problem given that bilateral flows influence each other and are therefore likely to confound clear understanding of migration. This thesis addresses these gaps by extending the standard gravity model in increasingly advanced ways to study China's interprovincial migration in 2010.

Interprovincial flows are not homogeneous and can be usefully classified into four different types: rural-rural, rural-urban, urban-urban and urban-rural migration. Results show that these four types of flow may differ from each other in their causes. Urban-urban migration is found to be under-researched but the most representative of all types based on number of similar coefficients. Therefore, it is then selected to further examine data dependence and non-linear distance decay. Results reveal that urban-urban interprovincial flows neither are independent nor entirely obey the gravitation law of distance decay.

In bridging these knowledge gaps, this thesis has made methodological and theoretical contributions to improving the understanding of China's interprovincial migration. The proposed multilevel gravity model could also be used in other bilateral flow studies such as trade and traffic, with particular strength in coping with data dependencies. My findings also have policy implications regarding migration governance and regional inequality reduction.

Acknowledgements

First, I would like to thank my supervisors Winnie Wang, Richard Harris and George Leckie. I would not be able to complete this thesis without their constant encouragement, guidance and support. They have made my PhD a fantastic experience and given me confidence in my own work. Winnie has given me the opportunity to do this PhD, and has been a wonderful role model and got me through the darkest moments of my life; Rich has patiently helped me to understand how to organise cluttered ideas into structured research, and has been giving me guidance in difficult situations; and George is always ready to help immediately and has not once complained at the constant stream of email questions that I send despite his own colossal workload, and has been giving me unwavering encouragement and support.

Next I must thank my proofreading team: Tim, Sarah, Gaz, Sam and Lucy. Thank you for always being there and undertaking marathon amount of proofreading for me! You have added a nice ‘English’ touch on my writing. I would also like to thank everyone in the Browns community, which is probably the friendliest environment in the world to do a PhD. My special thanks go to Emily and Claire for your effort in keeping me on the sanity side. I also want to thank Gwilym, Helen, Shawn, Alan, Amy, Beatriz, Stephanie, Helena, Diksha, Mike, Guillaume, Moya, Stefan, Chengxi and Tom in particular for bearing with my frequent tears and weird sense of humour. It has been a wonderful four years.

Thanks to my friends Lulu, Jing, Zhixin, Ersilia, Mary, Lei, Guanpeng, Chao, Wei, and Xiaojing for their ongoing support and encouragement. My family remains as the inspiration to conduct this research and deserves special thanks for giving me the courage to study abroad: other people in the village must have thought you went crazy to support me to travel this far from home. Mom, Dad and Bro: thank you!

Finally, I am extremely grateful to the Economic and Social Research Council for funding my PhD, to the Worldwide Universities Network and the International Organization for Migration for offering me the internship in Berlin to further my understanding of migration governance, to the Alumni Foundation for supporting me to present my research in China, and to Centre for multilevel modelling and Spatial modelling group for the stimulating academic environment.

Declaration

I declare that the work in this thesis was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the thesis are those of the author.

SIGNED: .... DATE:..........................

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

This chapter provides the overall research background for this thesis. It establishes the context, motivation, research questions, and how each chapter is structured to fit into the overall theme of this thesis.

1.1 Research background

Since the economic reform in 1978, China has witnessed large volumes of internal migrants. Based on the definition of Chinese Census 2000 and 2010[1], internal migrants refer to people who currently reside in one place but have their household registration in another county (city or district) and have left their registered residence for more than six months (Fan 2005a; Shen 2012). Because of its huge population size (1.3 billion in 2010, source: National Bureau of Statistics of China), the scale of such population movement is astonishing and often surpasses that of other countries. For instance, there were 294 million internal migrants in China in 2015 (Source: National Bureau of Statistics of China), which is 20% larger than the total number of international migrant stock of the same year in the world (Source: UN DESA). This massive internal migration is reported as primarily labour redistribution and economically-driven, which has been essential to China's rapid urbanisation and industrialisation (Fan 1997; Shen and Yang 2014).

The research topic of this thesis is the recent internal migration in China. This choice is motivated by the substantive, methodological and policy importance for doing the work. Research on internal migration in China has grown steadily over the past few decades but further advances in theory and the modelling strategies are still arguably in need. There remains a major limitation that persists throughout the established literature: although regional inequality has been put forward as an important driver of

[1] More detailed information of this definition (including its limitations and based on it how the migration counts are derived) is provided in Sub-section 3.1.1 of Chapter 3.

interprovincial migration, three major levels of regional inequality – rural and urban, province, and region have not been systematically investigated regarding their effects on interprovincial migration.

This limitation confounds the understanding of where, why and what type of migration is happening among scholars and policy makers. To illustrate, the recent converging provincial and inland-outland inequality might have incurred a decrease in interprovincial or between-regional movements, whilst the widening rural-urban inequality is likely to cause growth in migration between rural and urban areas (Herrerias 2012; Pannell 2012; Su, Liu et al. 2015). Therefore, this limitation inhibits a thorough understanding of the ways in which these three levels of regional inequality impact upon interprovincial flows, leading to misinformed policies of migration management and regional development. Indeed, policy makers may have directed resources in counterproductive ways to reduce regional inequality and improve migration governance, despite their continued endeavour to achieve both targets simultaneously.

Apart from the impressive scale, internal migration in China has other interesting features, including its unique state policies towards migration management and socio-economic development. China has achieved fast industrialisation and rapid economic growth, which is accompanied by massive internal population movement within the country (Lewis 1954a; Fan 1997; Fan 2005a, 2005b). Behind this massive internal migration, the major cause is economic growth but not vice versa (Zhang and Song 2003). The rapid economic growth is achieved by combining market forces and state interventions such as the implementation of a wide range of regional development policies (Lewis 1954b; Fan 2005a). Studying China's recent internal migration therefore also offers a valuable opportunity to further the understanding of migration in the context of a socialist transitional economy.

In China's context, the internal migration includes both interprovincial and intraprovincial flows. Currently there are 31 provincial administrative units (thereafter referred to as provinces) in mainland China. The National Bureau of Statistics of China makes this distinction in collecting the official census data. The difference between these two types of flows lies in whether the origin and the destination of a migration process are within the same provincial-level administrative unit. Therefore, between-province migration flows in general endure longer migratory distances than within-province flows. Based on the censuses of 2000 and 2010, within-province migration flows took a larger share (70.6% in 2000 and 67.1% in 2010) of the total nationwide internal migration than those of between-migration flows. However, rural migrants are still more likely than urban migrants to conduct interprovincial migration, although the share of between-province migration has decreased from 50.3% in 2010 to 46.8% in 2014 for rural migrants (Source: National Bureau of Statistics of China).

This thesis focuses on the research of interprovincial migration flows. This is because interprovincial migration flows better reflect the political and socio-economic connections between provincial administrative units on the nationwide level (Liang, Chen et al. 2002). For instance, directions of net interprovincial migration flows could represent the unbalanced socio-economic development between provinces (Fan 1995; Fan 2005a; Liu, Stillwell et al. 2014). By contrast, intraprovincial migration flows more reflect labour redistribution within provinces, primarily driven by intraprovincial factors such as county-level[2] economic growth (Zhang and Song 2003). Consensus in the literature is that China's state policies have greatly shaped the regional development landscape (Démurger, Sachs et al. 2002; Sun 2013; Liu, Stillwell et al. 2014). Here, 'region' refers to a group of provinces sharing similar state development policies and socio-economic features (Groenewold, Chen et al. 2010). Based on geographical coverage of major regional development policies, currently there are four economic regions in China: the East, the Central, the Northeast and the West. Regional development could affect how provinces exchange capital[3] and resources with each other. There is a strong association between regional development and interprovincial migration flows (Fan 1995; Fan 2005a, 2005b).

[2] A sub-provincial administrative unit in China.
[3] Capital here refers to its definition in classical economics as one of the three factors of production. The other two factors are land and labour.

Interprovincial migration flows within China are complex and dependent on many factors, meaning that modelling them accurately requires advanced methodological approaches. This is exemplified by the need to develop different extensions of the standard gravity model in this thesis to bridge specific knowledge gaps in studying China's interprovincial migration flows. These flows can be separated into different types based on their rural and urban components, and they have a unique structure and a possible non-constant rate of distance decay. For instance, interprovincial migration consists of bilateral pairs with possible strong correlations, and the adverse effect of distance upon migration is subject to the contiguity of provincial and regional boundaries. Each of factors deserves specific investigation and they altogether contribute to better understanding of associations between regional inequality and interprovincial flows. These literature gaps will be explicitly explained in next chapter.

1.2 Research questions

Regional inequality in China clearly has three major levels of regional inequality – rural and urban, province, and region (Fan and Sun 2008). The major motivation of examining the association between interprovincial migration and regional inequality lies in that these three levels have taken on different trajectories since the late 1970s. Studies show that inland-outland and rural-urban inequalities have continued widening overall, whilst provincial inequality has shown early signs of convergence after a long period of continuous growth (Herrerias 2012; Pannell 2012; Su, Liu et al. 2015). Moreover, major aspects of the regional inequality such as income, employment and access to public services have been extremely dynamic and rapidly changing over time and across locations (Jian, Sachs et al. 1996; Fan and Sun 2008; Fan, Kanbur et al. 2009).

Regional inequality has been put forward in the literature as a major driver of interprovincial migration in China (Fan 2005a; Shen 2016a). Regional inequality essentially affects the direction of net-migration between two regions, as more attractive one of the two regions pulls in stronger in-flows. This relationship could be better understood through a simplified model of a flow between two places. This flow is subject to combined influences of population, wealth, or other factors in the origin and the destination along with the origin-destination distance. Measurements of population, wealth and distance are indicators of prime importance to this migration flow. Increases in the population or wealth of one or both of the location will result in greater migration, whilst longer distance leads to less movement. Such associations can be expressed in equations analogical to Newton's law of gravity, with a migration flow treated as the attraction force between two places and population and wealth as their masses. This analytical framework is known as the gravity model of migration, characterised by proportional linkages between flow strength and migration predictors after logarithmic linearisation. Such proportional linkages are consequently interpreted as the gravitational law of migration. This thesis chooses the gravity model of migration as the theoretical and methodological framework because of its popularity in migration studies and flexibility in incorporating origin and destination factors to model migration flows (Peeters 2012; Flores, Zey et al. 2013; Fitzgerald, Leblang et al. 2014; Poprawe 2015; Shen 2015).

The overarching question of the thesis is 'what are the associations between China's interprovincial migration and three levels of regional inequality (rural and urban, province and region)?'. It is important to ask and answer this question, because former studies have not systematically investigated the association between China's interprovincial migration and all three levels of regional inequality. Such a limitation may not only have led to neglect of important rural and urban, provincial and regional nuances of substantive research interests, but also places studies at risk of a range of potential biases. In answering this research question, this thesis could therefore provide a richer understanding of China's internal migration by exploring these three major levels of regional inequality. With this increased knowledge, scholars will be better able to plan future work at understanding why migration flows are contextualised by different places in terms of rural and urban, provincial and regional levels, which will in turn enable policy makers to better target resources to reduce regional inequality and improve migration governance at the same time.

Within this central research question, three (sub-) research questions require answering to build up a coherent, reliable, and empirically informed argument:

(1) What are the associations between China's interprovincial migration and rural and urban level of regional inequality through origin and destination population, income and distance?

The first research question is important in examining how rural and urban areas are linked through migration flows between provinces. Such an investigation is conducted by using origin and destination population, income and distance. It aims to understand flow types flows (the rural-to-urban, urban-to-urban, rural-to-rural and urban-to-rural flows) through the lens of rural and urban inequality.

(2) What are the associations between China's interprovincial migration and province level of regional inequality through flow dependencies?

Flow dependencies derive from various correlations in the data, impacting interpretations of migration predictors. Although there are extensive studies of interprovincial migration flows, flow dependencies are inadequately examined. This research question essentially investigates province-specific effects on interprovincial migration flows, which is important for understanding associations between interprovincial migration and provincial inequality.

(3) What are the associations between China's interprovincial migration and region level of regional inequality through distance decay?

The third research question is to explore how regions are linked through migration flows between provinces. Such linkages are under influences of regional and provincial boundaries, particularly as the association between distance and migration is not always linear. It is therefore important for understanding how regions impact on interprovincial migration flows.

In answering these research questions, this thesis explores novel methods to extend the standard gravity model to study interprovincial migration in China. There are three major reasons motivating this thesis: first, the heterogeneity of interprovincial flows is under-researched as former studies have not systematically studied these four different types of flows (i.e. rural-to-urban, urban-to-urban, rural-to-rural and urban-to-rural flows), so the first research question needs to be asked and answered; secondly, there are various correlations in migration flow data, directly leading to the second research question; and thirdly, the rate of distance decay (here referring to the natural log of the migratory distance) can be non-linear and subject to the contiguity of provinces and regions, which is explicitly dealt with by the third research question. In bridging these knowledge gaps, this thesis proposes three different and increasingly advanced methodological frameworks in each results chapter. The next section will describe the thesis structure and all the chapters in more detail.

1.3 Thesis structure

This section gives an outline of the structure that the thesis will follow.

Chapter 1 Introduction

Chapter 1 provides the research background of this thesis. It discusses the research scope, motivation, and the main research questions of this thesis followed by an outline of the thesis structure and then a description of the nature of the thesis.

Chapter 2 Literature review

Chapter 2 reviews relevant migration theories and empirical studies, with special attention paid to China’s internal migration. In so doing, it lays the theoretical foundation for the thesis and identifies the knowledge gaps in previous work that this thesis sets to fill.

Chapter 3 Data and methodology

Chapter 3 details the development of measurements for predictors, demonstrating the rationale of using rural and urban segmented populations and incomes. It also explains how the standard gravity model comes into being in migration studies, laying the methodological basis for further extensions in the three results chapters of this thesis.

Chapter 4 Analysis of the rural and urban income divide and interprovincial migration in China from 2000 to 2010 with gravity models

This first research chapter presents the analysis of interprovincial migration in China from 2000 to 2010 with gravity models, answering the first research question. This chapter adopts a new approach that specifies the gravity model with rural and urban populations and incomes. The total migration flow between any two provinces is decomposed into four streams: urban-urban, urban-rural, rural-urban and rural-rural migration. By doing so, it is possible to estimate the contribution of the pull and push of both rural and urban areas across provinces for the years 2000 and 2010, and investigate the interprovincial migration flows of four different types in 2010.

Chapter 5 Analysing interprovincial urban migration flows in China: A new multilevel gravity model approach

This second research chapter proposes a multilevel extension to the standard gravity model to study the under-researched urban to urban migration in China, answering the second research question. Many previous studies have looked at rural to urban migration in the context of urbanisation and economic development, and at return migration. Very few have looked at what is becoming more important in increasingly urbanised countries, which is the movement from one urban location to another.

This chapter develops a new method that allows for the interconnections between migration flows: between those that share an origin, those that share a destination, and where there is a reciprocal flow between places. A conventional gravity model of migration ignores those connections, risking erroneous estimation of the regression parameters and of their statistical significance. It also ignores that those connections are of substantive interest – they reveal the interconnections between places regarding the numbers of migrants that they send and receive. This chapter motivates and illustrates the advantages of the new approach using urban-urban interprovincial migration with a sub-dataset.

Chapter 6 Exploring the non-linearity of distance decay in interprovincial urban-urban migration flows

This third research chapter answers the third research question, by exploring the non-linearity of distance decay with urban-urban interprovincial migration in 2010. Based on the multilevel gravity model in Chapter 5, this chapter strives to further extend the model by treating log distance as both linear and quadratic terms as well as adding new variables to investigate the contiguity of provinces and regions. This new approach has the potential to reveal the spatial distribution of migration flows in a more realistic way.

Chapter 7: Conclusion

This chapter summarises the main research findings, by addressing each of the research questions in order. It also details the main strengths and limitations of the data and methods utilised in the thesis. This chapter then argues for the contributions that the thesis has made to the academic world as well as the potential applications in practice such as policy making. Finally, this chapter points out directions of future substantive and methodological research that deserve further investigations.

1.4 The nature of the thesis

This section describes the nature of this thesis. Although the overall structure of the chapters described in the last section is in line with the organising principle of traditional theses, the main elements of this thesis (i.e. the three research chapters 4, 5 and 6) are presented as self-contained papers. As such, please note that there is some necessary repetition on description of data, context and theory in each of these chapters. More specifically, Chapter 4 and 6 are included as single-authored paper drafts, whilst Chapter 5 is a published paper with the title 'Analysing inter-provincial urban migration flows in China: A new multilevel gravity model approach' in Migration Studies (Zhang, Wang et al. 2018). I co-authored this publication with my supervisors Winnie Wang, Richard Harris and George Leckie, with me being the first author. In writing Chapter 5, I was originally modelling the cross-classified migration model using 'mixed' command in Stata, after having collected data, done the literature review and finished analysing the traditional linear regression formulation of the gravity model. George proposed me to move beyond it to the full model with the two additional correlations using 'runmlwin' command to call MLwiN in Stata and advised me to write the model syntax[4]. He also showed me how to implement this model in MLwiN with all the complicated constraints. I then re-analysed data before writing the discussion and conclusion of Chapter 5, and all my supervisors advised me to better present the results as well as helped me to improve the manuscript.

[4] The syntax can be found in Section Chapter 1 of Appendix.

Chapter 2 Literature review

This chapter reviews relevant literature in order to extend the theoretical and empirical background and identify knowledge gaps. The first section discusses recent developments and potential opportunities for new and innovative research in migration studies, by examining new theoretical trends, quantitative methods, data and ideas in studying China's interprovincial migration. The second section first describes the patterns and trends of internal migration for both developing and developed countries around the global, and then extensively discusses major macro- and micro-level migration drivers of internal migration in the world. The third section explains the scale and recent changes of internal migration in China first, and then moves on to discuss major macro- and micro-level drivers of China's internal migration. Also in the third section, this thesis focuses on explaining regional inequality as a major labour migration driver in China, and explicitly describes related policies as important background. The fourth section summarises how the new opportunities afforded by this methodology have allowed the development of each chapter for this thesis, which lays the foundations for a greater understanding of interprovincial migration in China.

2.1 Theories and methods in labour migration: opportunities for interprovincial migration in China?

This section attempts to explore the diverse nature of migration by reviewing the latest trends in both migration theories and empirical studies, with a focus on internal migration in China, seeking to integrate different strands of migration theory to enrich the theoretical framework for the thesis. The first sub-section will review different perspectives of migration theory, while the second sub-section will explain the new trend in labour migration studies.

2.1.1 Economic, socio-cultural and neo-Marxism perspectives of migration theory

There are three main perspectives of migration theory, namely economic, socio-cultural and neo-Marxism, on explaining the cause or determinant of migration (Ye, Wang et al. 2013; Shishehgar, Gholizadeh et al. 2015). As theoretical research goes through shifts in perspectives over time, from economic to neo-Marxism, with socio-cultural considerations in between, the following will explain these three perspectives in chronological order.

Economic perspective is the most fruitful one, which has produced a wide range of theoretical strands including both early and neoclassical economic approaches[5], dual labour market theory[6] and so on. The economic perspective of the migration theory can date back to Ravenstein's observations in 1880s (Ravenstein 1885; Ravenstein 1889), where he stated that economic factors are the major cause for most forms of migration (King 2012). Such an empirical generalisation reflects some crucial features of the neoclassical economics paradigm, which focuses on the analysis of labour mobility. Such analyses are prevailingly based on the assumption that individual migrants strive for the maximisation of utility and always make rational choice of destinations after comparing factor-price differentials across places (Borjas 1989). Previously, the neoclassical economics paradigm of the migration theory has been widely applied to explain migration on different scales (Massey, Arango et al. 1993). Many studies have endeavoured to use it to understand the macro-level spatial distribution of factors of production such as labour and capital or in the micro-level migration decision-making process (Sjaastad 1962; Borjas 1989; Massey, Arango et al. 1993). Indeed, the neoclassical economic approaches are very effective and useful

---

[5] Borjas (1989) summarised the neoclassical economic theory in migration studies as follows: 'Neoclassical economic theory assumes that individuals maximize utility: individuals "search" for the country of residence that maximizes their well-being' (Borjas 1989, p. 460-461). The neoclassical economic theory bases its analytical framework on the assumption of (potential) individual migrants being rational and always making choices to maximise the utility.

[6] Reich et al. (1973) explained the dual labour market theory as follows: 'The primary and secondary segments, to use the terminology of dual labour market theory, are differentiated mainly by stability characteristics. Primary jobs require and develop stable working habits; skills are often acquired on the job; wages are relatively high; and job ladders exist. Secondary jobs do not require and often discourage stable working habits; wages are low; turnover is high; and job ladders are few. Secondary jobs are mainly (though not exclusively) filled by minority workers, women, and youth' (Reich et al. 1973, p. 359-360).

in explaining migration decision-making on the micro level, where migration (who and when to migrate, and the choice of the destination) is the outcome of individual migrants rationally weighing up the pros and cons of all possible options with available information (Sjaastad 1962; Massey, Arango et al. 1993; King 2012). On the macro level, the theoretical hypothesis of the neoclassical economic approaches is that some regions have relatively plenty of labour but deficient employment opportunities whilst other regions have relatively plenty of employment opportunities but deficient labour, thus leading to the wage differences across different regions. Consequently, some labourers from low-wage places are attracted by (expected or real) higher economic returns in high-wage places and choose to relocate themselves there. Such relocation of labourers reflects and corresponds to the dynamics of supply and demand of labour across different regions on the macro level (King 2012). Compared with the micro-level application, the macro-level application of the neoclassical economics paradigm is therefore more relevant to this thesis, which focuses on the analysis of the interprovincial migration in China from the perspective of the uneven spatial distribution of labour and human capital return across urban and rural areas, different provinces, and different regions.

Under this perspective, economic factors are regarded as the main migration determinant. At the heart of many economic-perspective migration theories lies a common assumption of homo economicus[7], meaning that people homogeneously have good economic sense and act in their own self-interest.

[7] One exception is the microeconomic migration model, which is based on utility maximisation theory. Utility maximisation theory originates from Behavioural Economics and assumes that human has many biases and examples of irrational behaviour. Utility maximisation theory has been most useful for analysing long-run tendencies in the observed behaviour of individual migrant’s decision making. Empirically, it proceeds directly from objective ‘independent’ variables to migration behaviours - the act of moving and the destination choice of movers, with several forms of information entering the process including the range of alternative destinations available and their specific characteristics, present preferences and future outcomes (Goodman 1981; De Jong and Gardner 2013). The microeconomic model has evolved into a valuable analytic tool for the study of migration behaviours among individuals in terms of optimal information acquisition prior to migration and for generating empirically testable propositions (Goodman 1981). However, the general issue of perfect information assumption and the notion of optimality in utility maximisation theory has led some researchers to implicitly or explicitly question its reliability, as in reality a clear distinction can be drawn between the level and quality of information among individual migrants. For instance, few destination alternatives are taken into account by migrants before moving in reality, and migration decisions are often made with imperfect information, which all contrasts with the assumption of perfect information in the basic microeconomic migration model. Many scholars resort to search theory for resolution, which basically explores the notion of optimality in migrants’ decision-making by asking and answering two overriding questions - the extent to which it is due to lack of information if migration decision-making is not optimal and how migrants search information when it is imperfect (Molho 2013).

Among economic-perspective migration theories, Lewis's dual labour market theory, Harris and Todaro's (rural and urban) two-sector migration model (Harris and Todaro 1970) and Stark's relative deprivation theory of labour migration (Stark and Stark 1991) are important in furthering the understanding of migration through income disparity (Ye, Wang et al. 2013). Lewis (1954) argued that rural-urban income inequality triggers population transfer from agricultural to industrial sectors accompanied by the migration of the surplus labour force from rural to urban areas, and that this movement will not stop until the industrial sector fully absorbs the surplus labour. However, it is not necessarily actual income that encourages this labour transfer, as Todaro et al. (1981; 1983) pointed out. The expected higher income in the destination could stimulate the move from the origin for migrants, although some periods might see a high destination unemployment rate. However, the assumed income inequality of the origin plays a significant role as well (Stark and Taylor 1991). Indeed, there has been evidence that the likelihood of migration is positively related to the extent to which people feel deprived within the origin community (Stark and Taylor 1991). Equally important, the new home economics theory customarily views migration as part of the family's goal to maximise household income, which involves a conscious weighing of alternatives and a choice of the most preferred destination option available (Becker 1965; Ferber and Birnbaum 1977; Conway and Cohen 1998).

Other scholars argue that it is wrong to assume the linear linkage between (both real and expected) economic returns and migration, despite the continued popularity of treating economic returns as a major incentive for migration (Ye, Wang et al. 2013). This is because not the poorest areas have the highest migration rates (Connell, Dasgupta et al. 1976); the poorest are neither the most likely to migrate (Taylor, Arango et al. 1996), nor are they seeking for high economic returns but survival (Swain 1996). Indeed, economic-perspective theories cannot always fully explain migration decision-making, as migrants may not always making rational decisions to maximise utility. Therefore, there must be other determinants affecting migration apart from economic incentives. For instance, some scholars attributed migration to biased development policies and the lack of public goods provision in origin areas (Wang, Piesse et al. 2013), highlighting the over-simplified assumption of homo economicus in economic-perspective theories (Boswell 2008).

Socio-cultural perspective may offer alternative explanations on migrants' decision making, among which social network theory is the most popular due to its strong explaining power (De Haan 1999). Social network theory postulates that migrants are influenced largely by social network, previous migration experiences and social institutions (De Haan 1999). For instance, Boswell (2008) analysed the important ways how social ties influence migration, echoing Radu's (2008) argument that migration happens not in isolation but is influenced by one's peer group. In other words, externality in migration, such as social network, peer behaviour and herd influences, denotes important social influences, and plays significant roles in determining when and where migration happens. As Lee (1966) points out in the 'push and pull' theory, that migration in reality is hardly completely rational and many exceptions to the generalisation of migration could be found. In that sense, socio-cultural perspective is highly valuable by bringing social and cultural factors that have been neglected under the economic perspective into the migration analytical framework.

Theories of the socio-cultural perspective are established on assuming existence of some migration threshold, risking overemphasising uniqueness of every migrant by neglecting their commonality on the aggregate scale. This migration threshold measures a certain degree of the 'socio-cultural pressure' experienced by an individual, up to which no act is undertaken, but beyond which the current situation is recognised unsatisfying and this individual begins looking for move opportunities. Obviously, this migration threshold varies from individual to individual, highlighting unique socio-cultural incentives influencing each migration process. The uniqueness of the migration threshold arises out of the distinctive gap between the aspiration and the reality for individual migrants (Molho 2013). In attempting to alleviate or eliminate this aspiration-reality gap, individuals triggered by the threshold could conduct migration with little utility maximisation notion, which may be regarded as 'irrational' under the economic perspective. Indeed, it could explain quite well why migration happens in certain regions or among certain groups of people beyond some economic reasons (Clark 1982; Greenwood 1985; King 2011). Nevertheless, theories of the socio-cultural perspective do have some potential danger in overemphasising uniqueness of every migrant. For instance, theories of the socio-cultural perspective are quite plausible to explain subsequent or repeat migration on the individual level, but they do not work so well in analysing why migration occurs in the first place (Ye, Wang et al. 2013).

By contrast, theories of the neo-Marxism perspective strive to explain the advent of initial migration, by highlighting the necessity of incorporating power and structure into migration analysis. This perspective refers to a combination of thoughts from various schools seeking to broaden the Marxist theory, such as the world system theory and the cumulative causation theory (Massey, Arango et al. 1993; De Haas 2010a). To be more specific, it moves away from the traditional accusations of class warfare in orthodox/classical Marxism and stresses the monopolistic rather than the competitive nature of capitalism.

With the advancement of the decolonisation movement around the world after the Second World War, the role of neo-colonialism and corporate capitalism has been evolving and both have become essential to the analytical framework of the neo-Marxism perspective migration theories (Breman 1978; Bagnai 2009; King 2012; Molho 2013). For instance, there are strong international migration flows persisting between post-colonial societies and their former colonial countries, which is due to the pathway dependence of the pre-existing socio-economic and cultural ties such as administrative and linguistic links (Skeldon 2006; Berger 2009; Agnew, Mamadouh et al. 2015). Neo-colonialism is the last stage of imperialism, which is characterised by a new form of imperialism through exerting the great power of giant monopolies (Nkrumah 1967; Prahalad and Lieberthal 1998). Such a phenomenon is also known as corporate imperialism, or is more generally referred to as corporate capitalism (Wallerstein 1974; Prahalad and Lieberthal 1998). Corporate capitalism has systematically reshaped the political and economic structure both across countries and within nation states (Castles 2002; Skeldon 2006; Wang 2013). On the global scale, for instance, corporate capitalism has substantially facilitated the systemic stratification of countries by inevitably creating 'cores', 'semi-peripheries' and 'peripheries' (Wallerstein 1974; Froebel 1980). 'Core' countries consist of eminent developed capitalist powers such as the U.S. and Japan, which exert dominant economic and political forces upon 'semi-periphery' and 'periphery' countries through different degrees of asymmetric trade and capital flows (Wallerstein 1974; Froebel 1980; Bagnai 2009; King 2012). Understanding corporate capitalism is therefore vital to unpicking the mechanism and impact of international and internal migration under the current globalisation process (Froebel, Heinrichs et al. 1980; Cristobal-Campoamor 2009; Chan and Pun 2010; Chan 2010a; Davis, D'Odorico et al. 2013). Most importantly, China is often recognised as a typical 'semi-periphery' country and has been subject to the prevailing influences of corporate capitalism within this neo-Marxism theoretical framework (Francois 1994; Bagnai 2009; Chan 2010a; Chen, Ge et al. 2013). To illustrate, the onset of the impressive economic growth in China is believed to have a close connection with the penetration of corporate capitalism since 1980s under the disguise of foreign direct investment (Prahalad and Lieberthal 1998; Fujita and Hu 2001; Nee, Opper et al. 2007; Tsai 2007). The rising internal migration in China is thus often perceived as a response as well as a contributing factor to its economic growth (Fan 1995; Chan 2013; Ye, Wang et al. 2013). In this aspect, theories of the neo-Marxism perspective are very helpful in understanding the advent and primitive drivers of internal migration in China.

Under this perspective, migration is often rated as not a 'choice' for poor inhabitants in origin areas but rather the only option for survival (Breman 1978; Breman 1979; Bernstein 2010; Phelan, Nigam et al. 2011). That is because the means of subsistence have been largely commercialised under the overwhelming compulsion of economic forces in this contemporary capitalist world (Breman 1978; Breman 1979; Bernstein 2010; Phelan, Nigam et al. 2011). Therefore, depoliticised economic perspectives cannot explain either the forces of migration in the primitive capital accumulation periods or recent urbanisation processes where peasants become deeply absorbed into the capitalist market (Ye, Wang et al. 2013). Most importantly, the recent popularity of neo-Marxism analysis in migration studies implies an increasingly conflictual society in which classes are becoming highly differentiated.

Despite its variety of uses and applications in migration study, theories of the neo-Marxism perspective nevertheless have attracted three streams of criticism due to the often-misplaced overemphasis upon class. Firstly, theories of the neo-Marxism perspective tend to either overlook non-political-economic conflicts or attempt to describe them to be predominately political-economically defined in analysing migration (Arato 1993; Fan 1999). This could be inaccurate in many cases. Taking gender inequality in migration for an example, the fundamental aspect of gender conflicts is not primarily political-economic but more patriarchal (Rzhaniëtěsyna 1983; Fan and Huang 1998; Yeoh, Huang et al. 1999; Fan 2002; Fan and Li 2002; Fan 2003; Yang, Li et al. 2005). Secondly, subjectiveness tends to be overlooked under this perspective. This is because an individual might interpret his/her own class quite differently to what is objectively defined by the society. Such a self-identity crisis has abundant evidence in modern society (Alcoff 1988; Erikson 1994; Brubaker and Cooper 2000; Bottero 2004). Thirdly, it is problematic to assume that all capitalists are only profit-seeking. Indeed, capitalism may be economic in nature, but many people (some individual capitalists included) in a capitalist society[8] still aspire for equity and fairness in society as well as a sustainable capitalist economy (Featherstone 1988; Ikerd 2005; Streeck 2014). Therefore, both the second and third criticism contradict with the general assumption among many neo-Marxism researchers that the capitalist class rules and manipulates the whole society for their own benefits. This assumption also typically neglects that individual capitalists could deviate from collective properties of the capitalist class.

Overall, one converging trend in migration theories is quite pronounced and taking over, with different strands of theories influencing and incorporating with each other. To illustrate, all three perspectives point to the combined analytical framework of economic, socio-cultural, as well as political factors, as they are intertwined with each other and co-occur as determinants of migration. This is exemplified by the ever-expanding theoretical framework of the gravity model of migration, although relevant observations dated back to 1880s (Ravenstein 1885; Ravenstein 1889) and its original form primarily

[8] A capitalist society refers to a society that implements capitalism as a social order and way of life, by which this society functions on the accumulation and progress of private capital (Streeck 2014).

focused on explaining the Newtonian gravitational relationship between the movement of persons (or goods) and major predictors such as populations of the origin and the destination and the migratory distance (Zipf 1946). With the joint efforts from many scholars, the gravity model has experienced substantial improvement in both theoretical and mathematical frameworks (Wilson 1971; Flowerdew and Aitkin 1982; Willekens 1983; Alonso 1986). Indeed, it has been developed into a general schema by including economic, demographic, socio-cultural and economic-political migration determinants into its analytical focus (Vanderkamp 1977; Relethford 1986; Beine, Bertoli et al. 2014). A full discussion of the mathematical framework of the gravity model will be given in Chapter 3, along with the model derivatives that are adopted by this thesis.

Most importantly, labour migration has long been considered as the dominant trend (King 2012). For labour migrants, some determinants are dominant while others are peripheral, which vary greatly with different age groups of migrants and diverse households in specific regions or particular times (Ye, Wang et al. 2013). The next sub-section will devote to explaining how both new trends could offer new insights for studying China's interprovincial migration.

2.1.2 New theoretical trends and China's interprovincial migration

The theoretical convergence of migration theory has occurred across both international and internal migration but hold particularly strong in the study of internal migration, wherein the interruptions of transnational borders have limited impacts upon population movement within the state boundary. Based on different theories mentioned above, many quantitative approaches and models have made achievements in modelling and estimating migration flows in different contexts and scales (Schwind 1973; Lucas 2006; Portes 2007; Shi, Zheng et al. 2014; Yang, Han et al. 2014; Greenwood 2015; Huang, Li et al. 2015). Among these quantitative approaches and models, the gravity model of migration is widely regarded as a good tool and the most popular model analysing migration flows (Peeters 2012; Flores, Zey et al. 2013; Fitzgerald, Leblang et al. 2014; Poprawe 2015; Shen 2015).

Due to its comprehensiveness and inclusiveness, this thesis chooses the gravity model of migration as the theoretical and methodological framework to investigate China's interprovincial migration. This choice is closely related to the converging trend of combined analytical framework. Indeed, the gravity model of migration has important flexibility in incorporating origin and destination economic, socio-cultural, as well as political factors when the migration flow is modelled (Peeters 2012; Flores, Zey et al. 2013; Fitzgerald, Leblang et al. 2014; Poprawe 2015; Shen 2015).

This theoretical convergence of migration theory also points to the importance of labour migration in migration theoretical research (King 2012). Specifically, multiple strands of theories and frameworks have focused on the underlying causes of labour migration and treated it as an outcome of complex processes deeply conditioned by socio-economic and political contexts at different scales (De Haas 2010b). Labour migration theories tend to focus on the predominant role of income inequality in the labour migration process, although they attempt to examine migration in different perspectives, contexts and scales (Ye, Wang et al. 2013). Neoclassical economic approaches and the dual labour market theory, for example, put economic factors at the centre of the analysis framework as the main migration determinant. As for Lewis's unlimited supply of labour, Todaro's expected income and Stark's relative deprivation theories, they all build up on the common theme of income inequality.

While economic factors are deemed of prime importance among labour migration theories, there are contradictions between these theories in practical applications. The first case is the disputed role of income (or other equivalent measurements of the wealth) in the migration process. In the analytical frameworks of economic-perspective migration theories, rural-urban income inequality (Lewis 1954b), expected income in the destination (Todaro and Stilkind 1981; 1983), and relative deprivation in the origin (Stark and Taylor 1991) each plays the primary role in triggering the movement respectively.

This has led to different measurements of income (or wealth) in building models to estimate migration flows in specific practical applications, although income acts as one fundamental predictor in general.

In the context of internal labour migration in China, average income is seen as an efficient indicator of the overall income level for both rural and urban populations. For instance, Wang (2004) employed a choice-based model using provincial average household income per capita to analyse migration determinants. Some researchers choose to adopt provincial GDP per capita instead of income to explore spatial patterns of migration (Song and Wang 2005), whose preference for the former derives from the neoclassical regional growth and convergence theory (Barro and Sala-i-Martin 1990; Barro and Sala-i-Martin 1992a, 1992b). However, total GDP better represents the size of economy and (market and employment) markets and is therefore preferred to GDP per capita as suggested by neoclassical growth theory and human capital theory (Becker 1994; Aghion, Howitt et al. 1998). This is exemplified by a study of interprovincial migration in China, in which Wang, Wei et al. (2005) constructed a Cobb-Douglas production function[9] with total provincial GDP to analyse the macro-level influence of the interprovincial migration.

While all of these different measurements and model specifications of income levels have offered important insights into interprovincial migration in China, there has been a lack of investigation into the rural and urban components of provincial income. Multiple studies have found strong evidence in both developed and developing countries for links between rural-urban income divides and rural-urban migration (Zhao 1999; Seyfrit, Bjarnason et al. 2010; Mendola 2012; Villarreal and Hamilton 2012). However, this rural-urban income divide is conventionally interpreted as a one-dimensional measurement. The underlying assumption is that urban areas always remain favourable for internal migrants due to higher expected earnings (Ravenstein 1885; Petersen 1958; Berry 1976; Todaro and Stilkind 1981), ignoring the possibility that rural areas in one province can be more attractive than rural or even urban areas of another province. One major consequence of this is inadequate examination of

[9] The Cobb-Douglas production function is a particular form of the production function and is widely used to examine the relationship between the inputs (such as capital and labour) and the outputs. The Cobb-Douglas production function was developed by mathematician Charles Cobb and economist Paul Douglas in 1927 (Cobb and Douglas 1928), and is notably famous for being the first aggregate or economy-wide production function (Solow 1957).

destination choices in interprovincial migration flows. This is evidenced by the reality wherein a regional system has interprovincial migration flows in different directions, namely urban-urban, urban-rural, rural-urban and rural-rural streams. Thus, there is a need to draw on relevant labour migration theories to fully investigate the four types of interprovincial migration flows through rural and urban income components (Sheng 2011; Luo, Shen et al. 2014; Shen 2016b).

Secondly, flow data dependencies have been largely missing in the discussion of model development. Flow dependencies derive from various correlations in the migration flow data, which essentially render assumption of independence for linear regressions invalid and impact interpretations of migration predictors. There are various correlations in flow data, inevitably leading to different types of flow dependencies. To illustrate, origins or destinations are unlikely to be unique but rather necessarily repeated in the data matrix, directly creating group dependencies of origins or destinations. These two group effects may well be correlated, so are bilateral flow-pairs. All of these four flow dependencies have not been systematically examined, which leads to the potential for misinterpretations and estimation errors.

Flow dependencies have not been systematically examined in the literature, particularly dependencies of flow-pairs and between origins and destinations. A partial exception is a study in of the distances moved by residential migrants in England and Wales, which allowed for individual and contextual variations by origin and by destination (Thomas, Stillwell et al. 2015). Their study did not consider the correlations between either bilateral flow-pairs or origins and destinations of flows. Similarly, the relative emissivity of the origin and the relative attractiveness of the destination in the estimation errors have been measured using China's (Shen 2016a) and the U.S.'s internal flows (Chun 2008). Then again, dependencies of flow-pairs and between origins and destinations were treated as interaction effects between places in their studies.

The final theoretical dispute is the role of distance in the migration process, which is usually measured as the physical direct-line distance between the origin and destination (Beine, Bertoli et al. 2014). Theories of Zipf (1946) and Stewart (1960) postulate that the rate of total migration between two places is inversely proportional to migratory distance. By contrast, Stouffer (1940) posits that migration is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities. Despite different interpretations of distance, these theories have offered straightforward model specification methods through linear functions. Indeed, the linearisation of distance has historically received wide adoption in modelling migration flows in different contexts and scales (Stewart 1960; Wang 2004; Thomas, Stillwell et al. 2015), due to the relatively direct estimation process and the interpretation that it entails.

However, the linearisation of distance in the model specification does not consider situations in which non-linear relationships may apply (Stouffer 1940; Zipf 1946; Stewart 1960; Sjaastad 1962; Olsson 1965; Jiang, Wang et al. 2013; Molho 2013). The underlying assumption of the linearity function is a constant rate of distance decay (i.e. a proportional relationship between log distance and log migration), as suggested by the aforementioned labour migration theories (Stouffer 1940; Zipf 1946; Stewart 1960). In regional migration systems, however, contiguity between provinces and regions often matter, leading to the possibility of non-linear distance decay (Olsson 1965; Wolpert 1966; Jiang, Wang et al. 2013). This is particularly true in China's case, because Chinese provinces and regions are distinctive in their levels of development due to unbalanced regional development policies. Therefore, introducing a non-linear function of distance to explore contiguity between provinces and regional differences has the potential to estimate China's interprovincial migration flows in a more realistic way.

In summary, the impetus to examine China's interprovincial migration flows through rural and urban income components, quantifying flow dependencies and extending the specification of distance decay with a non-linear function can, however, only be tested and explored with appropriate data and measurements. Such data acquisition and measurement development process will be fully explained in the next chapter. The following section will focus on discussing the patterns, trends and major explanations of internal migration around the world, in order to understand the general characteristics and major drivers of internal migration in both developed and developing countries. The knowledge and learning from the next section therefore lays the foundation for situating China's internal migration in a global context.

2.2 Internal migration in the world: patterns, trends and major explanations

This section first describes the patterns, trends and major explanations in both developed and developing countries, and then moves on to the discussion of how China's internal migration links to and departures from the general characteristics and major drivers of internal migration around the world. Cautions should be taken that the general regularities are emphasised over the specifics in portraying the basic comparable contexts and settings of different countries around the world in an as clearly and succulent way as possible. This inevitably leads to certain reduction in attention paid to details. By doing so, this section aspires to conduct a systematic review of empirical studies of internal migration, in order to better understand China's internal migration from a global perspective.

2.2.1 Patterns and trends of internal migration around the globe: how developed and developing countries are different?

Empirical studies about internal migration are highly diverse around the world (De Haas 2007b; Molho 2013). In this sub-section, a general description of the macro-level patterns and trends of internal migration around the globe is conducted first. This sub-section then moves on to explain the distinguishing features of internal migration in developed and developing countries. The goal of this sub-section is to provide the background and context of how internal migration unfolds in different regions and countries, which will lay the foundation for the major explanations of why internal migration happens in the next sub-section.

2.2.1.1 General characteristics of internal migration around the globe

There is strong evidence observed around the globe that internal migration shares important similar interlinkages with the general social, economic, and demographic transformation processes across different countries (Zelinsky 1971; Wood 1982; De Haas 2007b; De Haas 2010a, 2010b; King 2012; Castles, De Haas et al. 2013; Skeldon 2014). This may indicate the existence of a fundamental socio-economic-demographic association underlying the complicated relationship between internal migration and development, although the historical conditions and geographical contexts under which internal migration develops are different across regions and countries (Zelinsky 1971; Wood 1982; King 2012; Castles, De Haas et al. 2013; Skeldon 2014). For instance, increase in internal migration rate is likely to be the outcome of development rather than under-development of a country, which points to a necessary condition for migration to happen – the possession of certain wealth among potential migrants to overcome the travelling costs and migration risks (De Haas 2007b).

Although the association between internal migration and development is complicated and varied over both time and settings, there appears to be an overall positive pattern around the world (Taylor, Arango et al. 1996). This is despite the (possible) brain drain caused by out-migration in less-developed origin places and the skills carried away by migrants, as well as any (potential) economic dependency and inequality increase in the origin induced by monetary transfers and other forms of migration remittances (Page and Plaza 2006; Portes 2007; Gamlen 2010; Greenwood 2015; McKenzie and Yang 2015; Ozden, Rapoport et al. 2015; Rezaei 2015). Conversely, transfers of human resources such as knowledge and skills that are carried by migrants, could also help to establish the mechanisms for liberalising the exchanges of goods, services and capital markets between the origin and destination places (Schwind 1973; Lucas 2006; De Haas 2009; Giralt 2015). Not only the positive association between migration and development seems valid both on a regional and national scale in the destination, but is an indispensable part of broader socio-economic change processes that are functionally and reciprocally connected with the development of the origin (De Haas 2007b). Bringing social and economic remittances together, it therefore could be argued that internal migration could facilitate poverty reduction and social welfare promotion in the origin as well as spur economic and social prosperity in the destination, if managed properly.

Apart from the observed complicated association between migration and development, some general demographic patterns and trends are evident among internal migrants in the world, despite regions and countries of considerable diversities due to different historical and geographical contexts. To be precise, in general migrants are more likely to be predominantly male (rather than female)[10], young (rather than old), single (rather than married) and better educated than the general population of the origin. This is because migration is a selective process in its dissemination and self-production among the general population; and migrants usually show mutiple distinctive demographic features compared with the general population as a result. In the established migration literature around the world, for instance, young people aged 15 to 29 tend to have a large share among migrants, due to their relatively lighter family obligations and higher employment aspirations (Findley 1977). Rapid population growth could also create employment and thus stand as a cause for out-migration (Zelinsky 1971; Findley 1977; Gedik 2005). In research literature of internal migration around the world, there is a general finding that male migrants tend to conduct pioneering or long-distance migration (Findley 1977), whilst female migrants tend to move with shorter migratory distances (Lucas 1997; Fan and Huang 1998). This may reflect that the traditional gender roles hold strong among internal migrants with more willingness instilled in males to take risks and explore (Eccles 1987; Oakley 2016), plus the fact that a large portion of female migrants are relocated for marriage purposes with relatively shorter distance from their origins (Fan and Huang 1998; Yeoh and Huang 1998; Eklund 2000).

Among all general demographic features of internal migration, the age selectivity is perhaps the most pronounced across different regions and countries around the world. Rogers and Castro (1981) drew on

[10] One exception is internal migration in a few countries of Latin America in 1970s, which contained more female migrants (Lucas 1978).

empirical studies of demographics and migration all over the world and summarised the fundamental regularities exhibited by migrants' age profiles with the Rogers-Castro curves[11]. The Rogers-Castro curve (Rogers and Castro 1981) obviously picks out the peak of migration in early adult years, whereas the simple human capital model does not. In other words, the narrowing life span to retirement may help to explain the diminishing rate of migration with age, but other factors must enter the story to gain a more realistic picture as to why migration at first rises than falls away very quickly after the early twenties - a pattern common to both developed and developing countries (Lucas 1997). The age selectivity could also reflect the impacts of family norms and life-course trajactories: at the certain stages in an individual's life, there is a high likelihood of changing households and communities at the same time (Li and Sologon 2014). Examples of such points in the lifecycle are when youth first leave the parental home in search of employment, upon marriage or divorce, and upon retirement (Findley 1977; Park and Kim 2015). In another recent example, Meeus (2012) found that the dissemination and reproduction of migration activities among the general population largely relied on whether individuals could respond to systemic and institutional changes through adjusting individuals' life courses in order to create favourable conditions for migration to occur.

There are strong general patterns of other demographic features of internal migrants observed around the world as well, such as education and employment. For instance, employment growth and demographic effectiveness were inversely correlated over the periods of 1980-81 to 1987-88 in the U.S, which resulted in changing spatial patterns of destination choice without changes in overall mobility rates[12] (Plane 1994). Another example comes from Gray and Mueller (2012), as they proved that crop

[11] There are three broad families of the Rogers-Castro curves under different retirement senarios (Rogers and Castro 1981). A standard Rogers-Castro curve consists of four major components (a single negative exponential curve of the pre-labour force ages, a left-skewed unimodal curve of the labour force ages, an almost bell-shaped curve of the post-labour force ages, and a constant curve), which can be expressed in a mathematical equation with a total of 11 parameters.

[12] There are actually two possible reasons for the inverse correlation between employment growth and demographic effectiveness as summarised by Plane (1994): either '(a) changing spatial patterns of destination choice without changes in overall mobility rates, or (b) lessened mobility while net population changes for states are maintained' (Plane 1994, p. 1554). However, the overall numbers of interstate migrants experienced small increases in both time periods of 1981-82 and 1986-87, as the absolute values of net migration increased across the entire set of states. This contradicts with the assumption in type (b) induced changes, which are characterised

failures and lack of local agriculture employment opportunities due to floods had a prevailing impact upon the growing out-migration rates in Bangladesh. Both cases may point to the same fact that migrants tend to have better education than those remain in the origin due to the selectiveness of migration (Findley 1977; Tharmaseelan, Inkson et al. 2010).

2.2.1.2 Internal migration around the globe: how developed and developing countries are different?

Internal migration in developed and developing countries has some structural differences. This is because countries from both groups tend to be in different evolving stages of the internal migration development. It is widely known in the literature the self-perpetuating and self-reproducing process of migratory behaviours among the population, indicating the mechanism of cumulative causation or chain migration (Myrdal 1957; Petras 1981; Massey, Arango et al. 1993). Because of this, migration itself is often seen as chain behaviour among the population and thus usually has a development course of its own (King 2012). In order to link the occurrence and evolvement of internal migration to broader demographic and development processes, this thesis adopts a diffusionist model of the 'transitional' perspective and concisely divide the development of internal migration into in three stages, namely the initialisation, the diffusion (or perpetuation) and the stabilisation (Zelinsky 1971; Wood 1982; De Haas 2007a; King 2012; Castles, De Haas et al. 2013; Skeldon 2014): the initialisation of migration refers to the onset stage of migration and features low and sporadic occurrences of pioneering migration, whilst the diffusion of migration is the self-perpetuating and self-reproducing process of migration among the general population through chain-causation and is characterised by rapid growth of mobility rate; and the perpetuation of migration could transition into the stabilisation stage through mobilising and re-distributing the factors of production (such as capital and labour) into equilibrium with a well-structured urban hierarchy, thus resulting in a stable mobility rate in a society (Zelinsky 1971; Wood 1982).

---

by diminishing total migration. Therefore, time periods of 1981-82 and 1986-87 were both featured with type (a) rather than type (b) interstate migration effectiveness increases.

Developed countries tend to maintain a stable mobility rate of internal migration with organised and ordered urban hierarchies upholding the distribution equilibrium of the factors of production (Zelinsky 1971; Wood 1982; Skeldon 2014). Some of the key demographic features are continuing low fertility and mortality rates that result in slight population increase (if any), with internal migration at stabilised rates being the primary source of population change (De Haas 2007a). Relevant evidence has been widely found in countries of Europe and North America (Zelinsky 1979; Van de Kaa 1987; Plane 1994; Wilson 2003; Raymer and Wiilekens 2008): the overall trend of stabilising mobility does apply to the developed world, although the exact demographic transition has shown evident diversities across countries of different historical and geographical settings. To elaborate, the stabilisation of migration takes place in the advanced and the future super advanced period of the famous five-stage mobility transition model[13] proposed by Zelinsky (1971). Based on the equilibrium model of migration, a society with stabilised migration has a well-structured urban hierarchy that could mobilise and re-distribute the factors of production such as labour and capital into equilibrium (Zelinsky 1971; Wood 1982).

Compared with internal migration of the developing world, internal migration of developed countries also tend to have distinctive compositional structures, with migration between urban areas taking the predominant role and population de-concentration from urban cores being the long-lasting trend (Zelinsky 1979; Fielding 1982; Champion 1989; Gedik 2005). The fundamental reason for the first compositional feature – the predominant role of urban-urban migration – lies in the relatively high and stable urbanisation rate in developed countries (Enyedi 1992; Kontuly and Geyer 2003; Kundu 2003). Everything else being equal, the larger the share of the urban population, the greater the likelihood of urban-urban migration taking place. The second compositional feature – the sustaining population de-concentration process from the urban cores – is widely known as counter-urbanisation, suburbanisation or urban decay (Berry 1980; Champion 1999). This population de-concentration trend is widely

[13] Zelinsky (1971) posited an idealised five-stage mobility transition model to explain the two closely related transitional sequences - the demographic and the mobility transitions, wherein the society passes through the premodern traditional, the early transitional, the late transitional, the advanced and the future super advanced period consecutively.

observed among developed countries. For instance, Fielding (1982) found strong evidence of counter-urbanisation in many cities of Western Europe. During the period of 1935 to 1980, migration systems in the U.S. also went through periods of con-urbanisation (namely urban region agglomeration) to those of population and economy de-concentration (i.e. urban core-periphery dispersal) (Plane 1984).

Developing countries are likely to have increasing mobility rates and to be in the initialisation and diffusion stage of migration, namely ranging from the pre-modern traditional to the early and the late transitional society (Zelinsky 1971; King 2012; Castles, De Haas et al. 2013; Skeldon 2014). The pre-modern traditional society is marked by little natural increase due to high fertility and mortality, the early transitional society is characterised by major population growth resulting from high fertility but rapid decline in mortality, and the late transitional society features significant but decelerating natural increase as a result of major decline in fertility with low mortality and large population base (Zelinsky 1971). Few contemporary developing countries are still in pre-modern traditional phase, whilst the majority are in the early or late transitional phase. Many demographic features of these mobility transition phases can be observed around the world in developing countries of Africa, Asia and South America (Skeldon 1992; Fields 1994; De Haas 2007a; Schiff and Ozden 2007; Bell and Muhidin 2009; Skeldon 2014).

Regarding the compositional structure of internal migration, developing countries tend to be undergoing the urbanisation process with rising share and importance of rural-urban migration (Johnson 1970; Renaud 1981; Bhattacharya 2002). The growing rural-urban migration has been greatly shaping the socio-economic outlook of the developing world with increasing forces (Todaro Michael 1976; Harris 1990; Lin 1994; Fan 2001). This is because contemporary developing countries in general tend to experience faster urbanisation and industrialisation processes than Northern European countries during the nineteenth and twentieth centuries (Kirk 1996).

This sub-section has provided a general description of the macro-level patterns and trends of internal migration around the globe and explained how internal migration of developed and developing countries is different from each other. Nevertheless, drivers of internal migration still need further exploration so as to understand the reasons behind these macro-level general characteristics and patterns observed around the world. The next sub-section therefore serves as an extensive discussion of different migration drivers for the internal migration around the world, in order to further explain how different migration drivers take effect in the migration process.

2.2.2 Major explanations of internal migration around the globe: how macro and micro factors are different?

Broadly speaking, there are five thematic categories of migration drivers: economic, political, demographic, social and environmental (Black, Adger et al. 2011). This sub-section examines these five categories of internal migration drivers at both macro- and micro-levels[14]. The reason of making this distinction of migration drivers is related to the fact that these five thematic categories of migration drivers are inter-linked to each other and co-affect the migration decision process. Therefore, even the same migration driver can operate and affect internal migration differently through distinctive pathways (direct, indirect or mediation, or confounding etc.) at different scales and levels (Rogers and Castro 1981; Massey, Arango et al. 1993; Finney and Simpson 2009; Nawrotzki and Bakhtsiyarava 2017). For instance, Gray and Mueller (2012) found out that the impact of disaster-induced crop failures upon

[14] Some researchers also proposed to add the concept of the meso-level migration drivers into the macro-micro binary system, such as social networks of individual migrants (Van Hear, Bakewell et al. 2018). This thesis does not adopt such a three-way categorising standard of migration drivers, as the meso-level migration drivers still fall into the distinction of internal-external agency of individual migrants. To elaborate, the conduct of migration activity is binary in real world – individuals will either conduct the movement or not; it is therefore more advantageous to adopt a binary categorising standard of macro- and micro-level migration drivers based on whether such divers are internal or external to individual agency of the migrant, when studying actual occurrences of migratory events as in the case of this thesis (Van Hear, Bakewell et al. 2018). Nevertheless, cautions should be taken that the micro-macro categorisation of migration drivers adopted in this thesis is to implement a structure for the purpose of the discussion rather than to suggest that micro- and macro-level migration drivers are isolated from each other or that one category of drivers are more important than others. In fact, macro- and micro-level drivers are interlinked and operate jointly to shape, enable or constrain migration processes.

mobility was actually in opposite directions on the household and local district scales in rural Bangladesh[15].

In this sub-section, macro-level drivers, such as political structures and barriers, urban structure and employment, and micro-level drivers of individual migrant's demographic features, life-course trajectories and family norms, will be extensively discussed in the following regarding how they drive internal migration to a greater and lesser extent in different parts of the world.

2.2.2.1 Macro-level drivers of internal migration: political structures and barriers, urban structure and demographic transformation

Following the established literature, this thesis refers to macro-level migration driver as factors that underpins the migratory decision making and drive the physical movement process but are external to individual migrants and their households and communities (Gardner 1981; De Haas 2011; Van Hear, Bakewell et al. 2018). Macro-level migration drivers usually include contextual migration determinants[16] of geographical, historical, institutional, economic and social dimensions in both migration sending and receiving areas and beyond (Gardner 1981; Carling and Collins 2018).

Macro-level migration drivers influence migration through complicated pathways. For instance, macro-level processes of the local, regional, national and even global scale, such as socio-economic

[15] Gray and Mueller (2012) noted that disaster-induced crop failures at household-level actually reduced mobility rate whilst only mobility of the sub-district level saw significant increase. Some possible explanations are that disasters can in fact increase local (especially the household) labour needs, and that disasters could also deprive local residents of the necessary means and resources to migrate out. On the sub-district level, however, mobility may still stand as an effective post-disaster coping strategy overall.

[16] This thesis focuses on studying actual interprovincial moves in China, it therefore does not make a distinction between migration drivers and determinants for the purpose of limiting the discussion to the outcome of physical population movement from the origin to the destination. In general terms, the migration driver is an analytical construct to understand how contextual factors influence the migration process, whilst the migration determinant directly affect the occurrence of migration (Carling and Collins 2018). Migration drivers and determinants are usually used as synonyms in the established literature. Nevertheless, such a differentiation between migration determinants and drivers may need to be considered in some cases where the development of migration aspiration and desire is of substantial research interest. This is because the aspiration and desire to move does not automatically result in a decision to move, nor does the decision to move always result in an actual move (Gardner 1981; De Haas 2011; Carling and Collins 2018).

development, demographic changes and political structures in origin and destination places affect migration within countries in different ways and to different extents (De Haas 2011). The conduct of a potential migration plan is unlikely to be independent of any macro-level migration determinants. For instance, the macro-level contextual factor of origin-destination inequality (here it means the difference of economic opportunities and social services between the origin and the destination) greatly influences how strong the migration aspiration of individuals or families is (VanWey 2005; Poon and Shang 2012; Villarreal and Hamilton 2012). The transportation infrastructure between the origin and destination directly determines what available facilities migrants could utilise in conducting migration (Cross 2001; VanWey 2005; Schiller and Caglar 2009; Poon and Shang 2012; Villarreal and Hamilton 2012). Another example is population density, which could reflect how fierce the competition for employment and social services in the origin and destination. Population density of the origin and destination therefore prompts potential migrants and migrant families to decide where to migrate through mediation pathways, as population density is closely related to migration-directly-relevant factors such as housing availability and employment rate in the origin and alternative destinations (Lee 1966; Gardner 1981; Goodman 1981; Greenwood 1993; Kofman 2004).

There are a wide range of important macro-level factors influencing the decision-making process of migrants. Migration decision making, according to De Jong and Gardner (2013), refers to the activity that the individual (usually within a family context) makes the decision to stay or how, when and where to move, based on evaluations of all the available information. As mentioned earlier, a wide range of factors jointly affect and influence the decision-making process and the conduct of migratory behaviours of migrants. Among them, some macro-level migration drivers have played a predominant role and have been most effective in mobilising the population and relocating them across different regions within many countries, such as socio-economic factors and institutional interventions (Fan 1996; Fan 1997; Fan 2005a, 2005b; De Haas 2011; Shen 2012; Carling and Collins 2018); by contrast, other macro-level drivers may have been unique and only influential in certain region of some country at a certain time – a perfect example would be an occurrence of a disaster, such as the Chernobyl disaster

(Kulakov, Sokur et al. 1993; Møller and Mousseau 2006). That is to say, some macro-level migration drivers appear to be more common and universal than others in many countries across the global (De Haas 2011; Lee, Carling et al. 2014; Carling and Collins 2018), all of which have undoubtedly influenced migrants and migration with varying effects upon population of different regions and countries (Ye, Wang et al. 2013; Otoiu 2014). In order to provide an analytical framework for drivers of internal migration in the world, this thesis therefore chooses to limit the discussion of macro-level factors to socio-economic factors and institutional interventions, due to their wide relevance and universal importance across different socio-economic, geographical, cultural and demographic settings (De Haas 2011; Lee, Carling et al. 2014; Carling and Collins 2018). As discussed earlier[17], there exists a universal socio-economic-demographic interlinkage underlying the internal migration system across different regions and countries (Zelinsky 1971; Wood 1982; King 2012; Castles, De Haas et al. 2013; Skeldon 2014). In other words, it would be difficult to exclusively examine the unique effect of one particular macro-level factor without considering its systemic and close association with other macro-level factors (Van Hear, Bakewell et al. 2018). Therefore, structural factors such as macro-level demographic transformations and socio-economic, cultural and political conditions (such as employment conditions and urban structures), will be jointly and extensively discussed in the following.

The urban hierarchy of a country reflects its urbanisation level and socio-economic development status (Brown and Moore 1970). Whilst urbanisation could facilitate the growth of cities by relocating some rural population into urban areas, Fielding (1982) noticed that the higher the rank of a city in the urban hierarchy, the weaker its population growth that was caused by internal migration in many cities of Western Europe. This strongly evidenced the process or the trend of counter-urbanisation. The structure of the urban system is also essential in redistributing population and other factors of production[18] on the aggregated level, which greatly facilitates the conditionalisation of the pathways of micro-level demographic factors such as individual demographic features and life-course trajectories, as well as

[17] See more detailed discussion in sub-section 2.2.1.1.
[18] See sub-section 2.2.1.2 for more details.

household features and family norms in the migration decision-making process for individual migrants and migrant households. For instance, some macro-level regularities of demographic and socio-economic characteristics of migrants could be observed in the stabilisation stage of migration (Lee 1966; Gardner 1981; Goodman 1981; Greenwood 1993; Kofman 2004): the difference of economic opportunities and social services between origin and destination could affect how strong the migration aspiration of individuals or families is to bridge the gap between the reality and their expectations, population density reflects how fierce the competition for employment and social services in the origin and the destination, and who and where to migrate is closely related to housing availability and unemployment rate in alternative destinations.

Migrants moving into metropolitan areas tended to have significantly different demographic and socio-economic characteristics with migrant conducting centrifugal movements from city centres in Western Europe (Fielding 1982). The correlation between migration and urban hierarchy is also observed in the U.S.. Plane (1984) examined the extent to which population movement efficiency affects population change in the U.S. for the period of 1935 to 1980, by describing the evolving patterns of population redistribution that can be separated into steady-state and non-steady-state patterns of population exchange[19]. During this period, migration systems in the U.S. went through periods of core region agglomeration to those of core-periphery dispersal and had thus brought substantial structural changes in the national urban hierarchy, as the U.S. economy evolved from the late stages of industrialism into the early phases of post-industrialism.

Such significant differences in the macro-level demographic and socio-economic characteristics of migrants conducting urbanisation and counter-urbanisation movements also point to an important

[19] Plane (1984) explained the difference between steady-state and non-steady-state patterns as follows: 'Steady-state patterns are conceived as those resulting from an existing arrangement of economic functions among centres. Non-steady-state ex-changes, by contrast, are those induced by changes in the location of economic activity, such as those associated with post-industrial structural change. Because steady-state flows result in little population change, whereas non-steady-state flows form the prime channels of interregional redistribution, demographic efficiency reflects the relative proportions of steady-state and non-steady-state movement' (Plane 1984, p. 295-296).

concept – the ‘escalator region’, which is featured with strong both in- and out-migration flows but with distinctive age structures and socio-economic statuses. In examining the link between social and geographical mobility in the U.K., Fielding (1992) defined an (upward) escalator region[20] based on three conditions: first, the regions would extend substantial attraction towards young people (particularly those with promotion potential) at the beginning of their career; second, the region would provide the necessary context for the young (both migrant and local) people to achieve accelerated upward social mobility with intra-regional labour and residential relocation movements; and finally, the region would export a significant proportion of labourers who are in their middle to later stages or their working lives. Although macro-level drivers for migration vary greatly across different geographic regions and countries, there is strong evidence for the existence of a positive relationship between social and geographical mobility among internal migrants observed in many ‘escalator regions’ of the world (Fielding 2010). Indeed, escalator regions tend to be more developed and offer more employment opportunities with higher returns to human capital and weaker institutional and cultural barriers for migrants (Fielding 1992). Conversely, a drastic rise in unemployment could greatly trigger out-migration (Van Hear, Bakewell et al. 2018).

Champion (1999) argued that urbanisation has been the single most important dimension of changes in population distribution at the global level, with the centrifugal population shifts of counter-urbanisation[21] playing an impressive and important role in shaping the population distribution across places. However, migrants conducting the urbanisation and counter-urbanisation movements tend to have distinctive demographic features and migration drivers: the urbanisation process is closely related to school leavers and young adults, who move to larger cities in seek of upward career and employment opportunities; by contrast, older people and families with children tend to move down the urban

[20] Fielding (1992) explained the concept of escalator region as follows: ‘It incorporates the idea that by migrating from one region to another, individuals ‘step off’ one escalator and ‘step onto’ another. From this perspective, the key to social promotion for the young and the ambitious is to combine an energetic pursuit of career advancement within their fields of expertise with a judicious use of spatial differences’ (Fielding 1992, p. 3).

[21] Counter-urbanisation refers to the process that larger cities lose population to smaller urban centres, as well as of a wider dispersal process that has produced a rural population turnaround (Champion 1999, p. 347).

hierarchy into less urban areas mostly searching for better retirement, housing and education amenities, who form the pillar of the counter-urbanisation movement. Based on analysing patterns of the net internal migration in Britain's 1991 Census, Champion and Atkins (1996) found a negative relationship between city's rank in the urban hierarchy and its rate of population growth induced by internal migration, which lent further support to Fielding's observation about the counter-urbanisation (Fielding 1982).

Population structure, socio-economic, cultural and political conditions of the origin and the destination on the aggregate level, such as employment conditions and urban structures, provide important contexts for the actual occurrence of migratory behaviour and the act of physical relocation process (Fielding 1982; Champion 1989; Fielding 1992; Plane 1994; Bell, Blake et al. 2002). High rates of rural population growth could often lead to rural-urban migration because the gap between population growth and the generation of employment opportunities is usually smaller in urban areas (Findley 1977). Conversely, the growth of non-agricultural industry and associated employment opportunities could reduce out-migration in rural areas. For instance, Wu and Yao (2003) showed that industrial structure change and the non-agricultural industry growth in rural areas had significantly affected migration decision-making among rural migrants.

Many macro-level structural factors of demographic transformation and socio-economic conditions (such as employment conditions and urban structures) tend to be relatively responsive to changes in policies of migration management (such as work-permit requirements or employment restrictions), which could play a decisive role in determining many important characteristics of internal migration (such as locations, volumes, directions, durations, types and forms). In fact, for most countries around the world state policies of internal migraiton management usually address three major issues – overpopulation and poverty in rural areas, rapid expansion of metropolitan areas, and unbalanced regional development (Findley 1977). The purpose of many policy interventions of internal migration is to optimise development by reducing migration pressure whilst widely spreading opportunities of social mobility among the general population (Van Hear, Bakewell et al. 2018).

Policies and institutional practices in other aspects of socio-economic management – such as trade and industry development, education and research, social welfare and infrastructure investment – are also likely to be migration drivers enabling or constraining internal population movements to different extents in different regions and countries (Van Hear, Bakewell et al. 2018). It also means that even socio-economic structures of population movements are politically constructed and constrained. It is therefore important to view various aspects of migratory processes and understand their macro-level institutional drivers through the lens of the underlying universal socio-economic-demographic association across countries. By jointly assessing macro-level migration drivers of demographic features and socio-economic, cultural and political conditions, it is now possible to explain the macro-level drivers for the major mechanisms of internal migration around the world, and to examine on the aggregated level the implications of internal migration beyond what it means for individual migrants, households, communities and societies that migrants move away from and arrive at within different countries.

2.2.2.2 Micro-level drivers of internal migration: demographics, life-course trajectories and family norms

Whilst the last sub-section has explained how major macro-level factors shape opportunity structures and condition people's migratory behaviours by simultaneously enabling and constraining the occurrence of internal migration, this sub-section now moves on to discuss how micro-level drivers of internal migration operate and take effect under different geographical and historical, political and cultural, and socio-economic settings. Micro-level migration drivers are traditionally defined as a combination of interconnected structural elements of socio-economic, cultural and political factors that are external to human agency and have a direct behavioural link to individuals by simultaneously enabling and constraining the occurrence of migration events on the micro level (De Haas 2011; Castles, De Haas et al. 2013; Carling and Collins 2018). It has been recently proposed to expand such a definition to include human agency so as to better describe and facilitate the research on the causality link running through people's agency to the occurrence of migration events (De Haas 2011; Carling and Collins 2018). Indeed, human's decisions are subjective to internal emotions, imaginations, aspirations, desires and preferences, although such decision-making is traditionally viewed as to have a dimension of rationality and is constrained by structurally determined macro-level resources and conditions (De Haas 2011; Carling and Collins 2018). By recognising people's internal desire and behavioural endeavour to change social structures (De Haas 2011), this thesis treats migration aspirations and propensities as dependent upon migrants' agency and endogenous elements of migration dynamics. In doing so, this thesis also maintains the logic consistency in adopting the binary categorisation of macro-micro migration drivers.

With regard to the migration decision-making process, a wide range of migratory determinants serve as important migration drivers. In order to better understand the mechanism of the migration decision-making process, this thesis adopts an established approach in the literature to decompose the whole migration decision-making process into three stages, namely when migrants first decide to leave their origin (Stage 1), when migrants start to search for destinations (Stage 2), and Stage 3 when migrants finally choose among all the alternative destinations[22] (Rossi and Shlay 1982). It then becomes clear that migration determinants of the individual and household level may maintain the direct and prevailing explanatory power towards the conduct of decision-making for the individual or the household, including demographical factors and life-course trajectories of the migrants, as well as household

[22] Based on migration theories of the socio-cultural perspective (see Sub-section 2.1.1 for more details), individuals of the population are all subject to the 'socio-cultural pressure' that represents the gap between the expectation and the reality of life quality (Molho 2013). When the 'socio-cultural pressure' exceeds certain threshold of some individual, that individual will deem the current situation in the origin as unsatisfying and will aspire to leave the current location, thus initiating Stage 1 of the migration decision-making process. In the subsequent Stage 2, the individual will actively start to collect and search for information of potential destinations. Once the individual finishes the information collection, he or she will analyse the information by weighing up the expected costs and returns of the move and come up with a (or a group of) desirable destination(s) from all possible alternatives (Stage 3 of migration decision-making).

features and family norms of the migrant households (Rogers and Castro 1981; Massey, Arango et al. 1993; Finney and Simpson 2009). This is because many demographic characteristics - such as age, the extent of schooling and marital status - are endogenously determined together with the migration decision (Lucas 1997). Taking education as an example, it can usually instill an awareness of and desire for better life opportunities in people. The educated among the general population therefore are more likely to aspire to maximising their returns, which stands as the internal driving force in individuals to choose to migrate (Findley 1977; Solis, Pullum et al. 2007). Education also has an important impact upon migrants' accessibility to employment information and opportunities as well as economic returns of human capital. Hence the general trend is observed around the world that migrants tend to have better education than those remain in the origin due to the selectiveness of migration (Findley 1977; Tharmaseelan, Inkson et al. 2010). In other cercumstances, education can also be an important driver for family migration in search for better future for the children (Findley 1977).

Family and social norms and expectations are related to age, life stages or life-course trajectories, whilst social transitions are important contexts for the effects of family and social norms and expectations upon migration to take place (Findley 1977; Mortimer and Shanahan 2007). To illustrate, an individual's migration motivation is influenced by the structure and function of the household and community (Yorimitsu 1985; Greenwood 1993; Yang and Guo 1999), and this process is taken place through social networks and emotional ties (Massey, Arango et al. 1993; Molho 2013; Otoiu 2014). Lee (1966) argued that individuals are subjected to the influences of household and community socio-economic characteristics, which have shaped their locational perceptions and preferences to a large extent. To elaborate, differences in socio-economic characteristics within household and community could have a significant impact on individuals' access to information. This is extremely important as it basically determines how much knowledge and control the migrant has over risks involved in the migration process. In that sense, migration could often be a joint decision within the household, which is subject to the influences of family norms and could be a calculated strategy of family risk spreading (Lucas 1997). Zelinsky (1971) explained the close relationship between lifetime cycle of residential shifts and schedules of circulatory migratory trips[23]. Based on life-cycle theory, migration occurs at certain stages in an individual's life trajectory to fulfil both migrants' and their families' certain needs at certain time (Findley 1977). The theoretical literature on family strategies - linking migration with such issues as fertility, education, marriage, inheritance, and risk spreading - has proved very fruitful in the last few years.

Social descrimination due to race, language, religion or places of origion is an important factor affecting migration decision-making and could fundamentally influcence where migrants choose to work and live. Finney and Simpson (2009) elabourated on the relationship between race and migration in the U.K., and showed that levels of xenophobia and racism, history of immigarion, housing prices and availability of socio-cultural support might be reasons why some migrants of minority groups chose certain neighbourhoods and communities to work and live. Similarly, Maya-Jariego and Armitage (2007) argued that the experience of migration could limit the sense and participation of community for migrants of multiple locations, echoing the finding of Crivello (2015) in exploring the role and position of migration in shaping young Peruvians' future of imagination.

Subjective emotions, imaginations, aspirations, desires and preferences of individuals are inseparable from the process of migration decision-making and the conduct of movement (Carling and Collins 2018). By bringing human agency and subjectiveness into the analytical framework, the complex realities of internal migration can be better examined. Indeed, such micro-level drivers of human agency and subjectiveness are deeply embedded in individuals' daily life experiences and social relations, and more broadly constrained by socio-economic settings and institutional regulations (Carling and Collins 2018; Van Hear, Bakewell et al. 2018). The micro-level migration drivers of human agency and subjectiveness therefore play a fundamental role of linking macro-level contextual influences to micro-

[23] Zelinsky (1971) summarised it as follows: 'Starting with one or more migrations to prep school or college, there follow movements incident on military service, marriage, job assignments either from one employer to another or within a single corporate structure, and, finally, to place of retirement. All this is aside from brief circulatory excursions of a few days' or weeks' duration' (Zelinsky 1971, p. 216).

level individual behaviours (De Haas 2011). In the latest literature, migrants are increasingly viewed beyond the traditional perspective of economic rationality, and rightfully recognised as autonomous agents who within their power can actively take measures to control and shape migration processes and outcomes. For instance, the traditional forced–voluntary dichotomy of migration is challenged and reconstructed by humanising the migration experiences of Afghan and Pakistani migrants to Europe (Erdal and Oeppen 2018). Subjective human emotions are also important to investigate the identity, sense of belonging, as well as other lived experiences such as integration in migration research (Boccagni and Baldassar 2015).

Each of these micro-level migration drivers has been crucial to unpick the causal mechanisms and pathways underlying the occurrence of migration events. They have also linked the micro-level migratory behaviours to broader socio-economic and political processes through a structure-agency framework (De Haas 2011; Carling and Collins 2018). That is to say, macro- and micro-level drivers are interlinked and operate jointly to shape, enable or constrain migration processes. For example, gender functions simultaneously on both micro- and macro-levels, with evidence showing that both macro-level cultural constructs of traditional gender roles and micro-level household power structures exert strong influences on female rather than male migration by limiting female migrants' participation in economic activities (Fan and Huang 1998; Yang and Guo 1999; Fan and Li 2002; Kofman 2004; King and Skeldon 2010). It is therefore extremely important to consider the joint influences of micro- and macro-level drivers in studying migration mechanisms and processes.

2.3 Internal migration in China: characteristics, recent changes and drivers

The last section has extensively discussed internal migration around the global in terms of its patterns, trends and major explanations, which provides important learning and knowledge for studying China's internal migration. By systematically reviewing internal migration in the world, the distinctiveness of

China's internal migration and the economic focus of this thesis can be better contextualised and acknowledged in the following.

This section will focus on discussing the scale and distinctiveness of Chinese migration and recent changes when it is compared to other modernising countries. The following section will then endeavour to explain major drivers of the internal migration in China, aiming to clarify why this thesis chooses to treat regional inequality as a major driver of labour migration in China from dimensions of rural and urban divide and unbalanced regional development policies so as to address the three research questions - (1) What are the associations between China's interprovincial migration and rural and urban level of regional inequality through origin and destination population, income and distance? (2) What are the associations between China's interprovincial migration and province level of regional inequality through flow dependencies? (3) What are the associations between China's interprovincial migration and region level of regional inequality through distance decay?

2.3.1 Characteristics and recent changes of internal migration in China: how is China distinctive?

This sub-section will first explain the major characteristics of internal migration in China, and then moves on to discuss political dimensions of rural and urban divide and regional inequality and how they are relevant to internal migration. This sub-section will end with the examination of the recent changes of China's internal migration.

2.3.1.1 Characteristics of China's internal migration

Compared with other modernising countries around the world, China's recent internal migration story is unique for two reasons: first, the state keeps relatively strict and effective management on the migration and mobility of its citizens through multiple policies; second, China's internal migration forms one of the largest 'temporary' population movements in human history with a peak volume of

298 million in 2015[25]. The temporariness of China's internal migration lies in its extensive institutional interventions (Liang and White 1997; Fan 2007); so much so that migrants have to adopt circular migratory strategies (Liang and White 1997; Zhao 2003). Also due to the temporariness of China's internal migration, measurements of internal migration have been notoriously inconsistent and difficult to detect to the extent of being 'statistically invisible' in the Chinese data (Fielding 2010; Chan 2013). This makes the internal migration in China a remarkable phenomenon but rather complicated to study.

For most countries around the world, nationwide internal-migraiton-related policies usually addree three major issues – overpopulation and poverty in rural areas, rapid expansion of metropolitan areas, and unbalanced regional development (Findley 1977). In the context of China, it has the one-child policy (recently abolished at the end of 2015) coping with overpopulation (Greenhalgh 2008), the infamous Hukou system (also known as the household registration system) designed for managing the speed of urbanisation and the expansion of metropolitan areas (Chan and Zhang 1999; Chan 2009; Bao, Bodvarsson et al. 2011), and urban-preference and unbalanced regional development policies to cooridinate the pace of regional development with the dynamics of spatial distribution of labour (Bao, Chang et al. 2002; Fan 2005a; Fielding 2010; Chen and Groenewold 2011).

Despite these macro-level nationwide internal-migraiton-related policies, China also has numerous regional and provincial internal-migraiton-related policies tailored to more nuanced local socio-economic development needs. For intance, first-tier mega cities such as Beijing and Shanghai have the most strigent Hukou policies in China, which only allows internal migrants with specifically needed qualifications and skills to obtain the local Hukou to permantly settle down (Wang and Zuo 1999; Guo and Iredale 2004). As Goldstein and Goldstein (1987) summarised, China's state and urban policies had been primarily oriented towards the management of internal migrants, and it had led to a striking contrast between the designed effects of such polices and their actual achieved results. This also have created tremendous pressure in simutaneously achieving development and migration control for many local government bodies, which have subtantially increased the risk of bad governance practices by accidently implementing contradictory policies. For instance, in late 1970s China adopted a nationwide policy of promoting the development of small towns and rural enterprises in order to reduce migration (Byrd, Byrd et al. 1990; Liang and White 1997); after a few years, many of the rural areas with successful rural enterprises did see evident decrease in out-migration, whereas they also experienced more significant increase in in-migration (Ma and Fan 1994; Liang and White 1997). In the end, despite being very effective in promoting socio-economic development (Ma and Fan 1994), such a policy was greatly revised in 1993 due to its very limited success in reducing total migration (Liang and White 1997).

As shown above, politics and the state are essential to understanding the characteristics of China's internal migration (Brettell and Hollifield 2014), the following two sub-sections will devote to discussing major state policies of rural and urban and regional levels in more detail. By doing so, the research context can be better depicted regarding the political and socio-economic conditions in China.

2.3.1.2 Rural and urban divide and migration policies

Migration and development are managed by the state government in China but faced with increasing market-oriented forces (Young 2013). Chinese government has a long tradition to control population with policies such as the Hukou (户口 *hukou*) system and the one child policy (Goldstein and Goldstein 1987; Young 2013). Under these policies, migration flows of certain direction such as the in-migration to small and some medium cities are favoured whilst those of other directions such as the in-migration to big cities are discouraged (Chen 1991; Lu and Wan 2014; You and Yang 2017). Migration management policies could therefore directly shape the volume, dynamics and geographical patterns of internal migration flows in China. Although this thesis does not focus on directly measuring influences of migration policies due to data limitations, it is still important to understand how migration flows are managed.

Among all the migration management policies, the Hukou system is the most significant and widely accepted as the root of rural and urban divide (Liu 2005). The Hukou system is a defining factor to classify interprovincial migrants into different types in this thesis based on their Hukou statuses. The current form of Hukou system came into effect in 1958, but with its origins dating back to ancient times (Young 2013). It is designed to restrict and regulate population movements by assigning migration quota through managing the admission systems, regulations of entry, duration of stay, work permits, and the access to social welfare support (Figure 2.1).

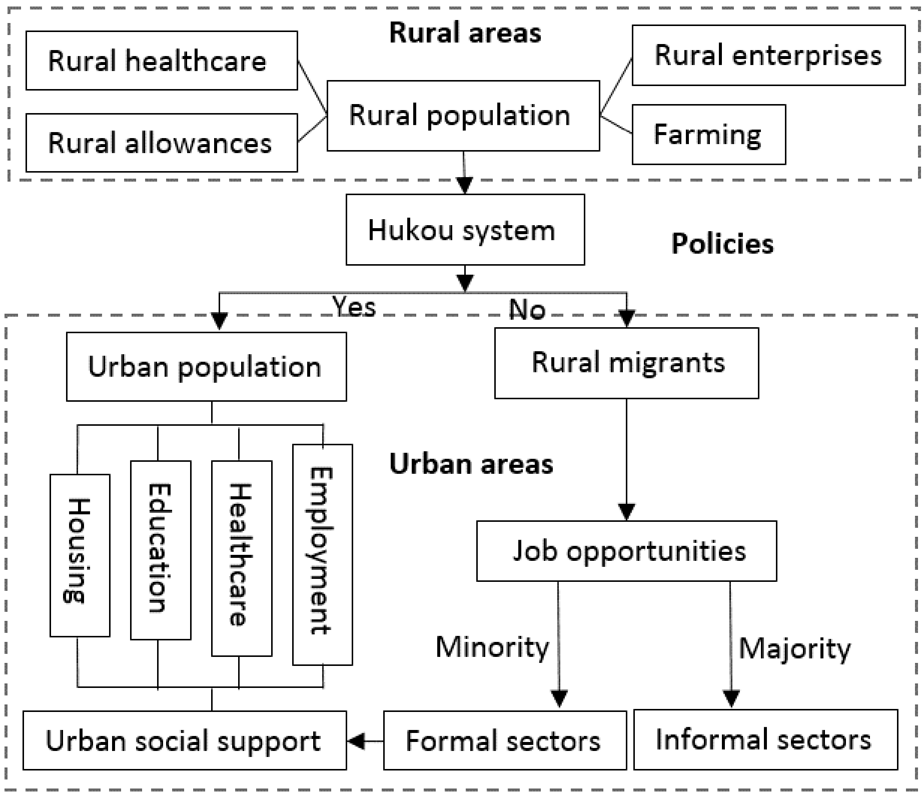

Figure 2.1 Hukou system and rural-urban migrants

This Hukou system has had great impacts on China's internal migration and mobility (Young 2013). In this Hukou system, the government has the information of the residential registration place 'urban households' or 'rural households' for the citizens, and every status holder is either recognised as 'agricultural' or 'non-agricultural'. In particular, the labour force transfer from rural to urban areas has been faced with many constraints from the Hukou system (Vendryes 2011; Wang 2012), which allows urban residents to have privileges on employment and other public services in protecting urban welfare system as shown by Figure 2.1.

The Hukou system also divides China into rural and urban areas, where land is either owned by rural committees or city councils (Chen 2011; Chen, Ye et al. 2014). In rural areas, farmers could get rural allowance from the government in subsiding farming activities and daily life since 2006 (Cheng 1991; Becker 2003; Bagnai 2009; Bao, Bodvarsson et al. 2011; Che 2014; Long 2014). However, a worker of rural origin seeking non-agricultural employment in urban areas is required to apply through the Hukou system via relevant bureaucracies to get the work and resident permit (Figure 2.1). Unless granted the access to the urban Hukou system, this rural worker is not entitled to urban welfare system (including employment, healthcare, education, housing, etc.). This rural migrant worker therefore might be confronted with the loss of both rural and urban social welfare supports (Liu 2005; Kuang and Liu 2012; Tao, Hui et al. 2015; Wang, Guo et al. 2015a). In short, rural migrant workers encounter all kinds of discrimination when they migrate to urban areas (Kuang and Liu 2012). By contrast, there is less restriction upon migrants of urban origins who are seeking employment in other cities, despite other demographic advantages that urban migrants have such as higher education and greater urban cultural and social capital (Liu 2005). It is relatively easier to move within urban or rural areas respectively as evidenced by urban-urban skilled-labour (Fan and Huang 1998; Liu and Shen 2014a) and rural-rural marriage migration (Fan and Huang 1998; Liu and Shen 2014b), whilst moving between rural and urban areas is faced with stronger institutional constraints from the Hukou system (Byrd, Byrd et al. 1990; Fan, Kanbur et al. 2011; Young 2013).

The role of the Hukou system has been evolving. Important reform in 2000 allowed rural migrant workers employed in formal sectors to have basic social welfare and support either in the origin or the destination (Figure 2.1), but it still acts as an institutional barrier to population mobilisation and migration overall (Chan 2010b; Wang and Fan 2012). Nevertheless, surplus agricultural workers continue to migrate despite this, due to the stimulation and facilitation of the industrialisation and the urbanisation process occurring across China in recent decades (Bao, Bodvarsson et al. 2011; Bosker, Brakman et al. 2012; Afridi, Li et al. 2015). This has forced the government to make further changes to population management such as the recent abolishment of one child policy and the loosening of Hukou control.

2.3.1.3 Regional inequality and unbalanced regional development policies

During the past four decades, China has experienced considerable socio-economic transformation in which the regional development policies have played an important role. The strong association between economic development and policies is considered as the main driving force of China’s rapid industrialisation and urbanisation process (Fan 2005a, 2005b; Shen 2013; Sun 2013; Liu, Stillwell et al. 2014). This sub-section examines the relationship between regional economic development and policies, providing context for recent interprovincial population movement.

Since the establishment of the People’s Republic of China in 1949, its regional discriminative policies have gone through five stages in general (Sun 2013). Stage one lasted from 1949 to 1978 and involved almost 30 years of central planning, which resulted in the overly agriculture-based economic structure before the economic reform in 1978. Stage two spanned from 1979 to the early 1980s and was characterised by a clear coastal-region-oriented development strategy with limited trade freedom in Guangdong and Fujian of the east region. The next stage was between 1985 and 1989, with enhancement of the coast-favoured strategy by widening free trade rights and regions. 1990s marked the arrival of stage four, emphasising the extension of free trade rights to several inland cities along the Yangtze River. Post-2000 has seen the emergence of stage five, when the whole of China was opened up with three major economic region policies - China Western Development, Revitalise Northeast China, and Rise of Central China.

These three major economic region policies extended the regional policies such as free trade rights and tax reductions that had been concentrated on major cities of the East to all levels of administrative units in the Central, Northeast and West regions (Table 2.1). The National Development and Reform Commission ('NDRC') has overseen creating and implementing regional development policies since 2003 (Sun 2013). The NDRC regional policies usually provide favourable tax reduction and state support for certain industries in certain regions (Groenewold, Chen et al. 2010; Sun 2013). For example, China Western Development was designed to promote socio-economic development in the western provinces to fix adverse impacts of former "uneven regional development strategies" by offering reduction or remission of taxes in some industries (Sun 2013). The Central and Northeast regions also got similar policy support through Revitalise Northeast China in 2003 and Rise of Central China in 2004. To mark the importance of these three policies, the NDRC set up special departments to map up region-specific policies and manage the implementation of these policies.

Table 2.1 Timeline of China's economic region policies and the respective departments in NDRC

Year	Policy	NDRC Department	Number of provincial units	Location
2000	China Western Development	Department of Western Region Development	12	Chongqing, Shaanxi, Sichuan, Gansu, Qinghai, Yunnan, Guizhou, Guangxi, Inner Mongolia, Ningxia, Xinjiang, and Tibet
2003	Revitalise Northeast China	Department of Northeast Regional Revitalization	3	Heilongjiang, Jilin, and Liaoning
2004	Rise of Central China	Department of Regional Economy	6	Henan, Hubei, Hunan, Jiangxi, Anhui and Shanxi

The scale and direction of internal migration has reflected the regional development landscape and how it responds to the regional development policies (Fan and Sun 2008; Ye, Wang et al. 2013). Empirical evidence from international settings has demonstrated a strong association between social-economic inequalities and migration flows (Vignoli 2008; De Haas 2010b; Ye, Wang et al. 2013). In a wide range of migration theories, regional socio-economic inequalities are usually regarded as a strong trigger: Lewis's unlimited supply of labour (Lewis 1954b), Harris and Todaro's expected income (Harris and Todaro 1970) and Stark's relative deprivation (Stark and Taylor 1991) all have stressed the predominant role of the regional income disparity in different contexts (Ye, Wang et al. 2013).

There is a continual need of state policies to promote the transfers of resources such as knowledge and skills and to liberalise the exchanges of goods, services and capital markets among different regions (De Haas 2009). The impacts of development policies however vary greatly with different groups of migrants from diverse backgrounds in specific regions and at particular times (Ye, Wang et al. 2013). Specifically, internal migration within the borders of a specific country has a close relationship with regional inequality. Unlike international migration, internal migration is more accessible for potential migrants and less likely to be rated as a threat or burden to other nations' economic growth and welfare systems, since migrants conducting the internal migration do not need to trespass borders or overcome socio-cultural barriers (De Haas 2007a; Ye, Wang et al. 2013).

However, inequality between spatially distinct areas within a country is still a major issue for all countries and their policy makers (Fan, Kanbur et al. 2009). China, particularly, has been experiencing severe regional inequalities for many years (Jian, Sachs et al. 1996; Fan and Sun 2008; Fan, Kanbur et al. 2009; Wang, Piesse et al. 2013). Although the Chinese government has made great efforts to address this issue, regional inequality has remained for decades (Knight and Song 1999; Fan, Kanbur et al. 2009; Wang, Piesse et al. 2013; Lai 2014). Moreover, regional inequality in China has been widely recognised as an effective predicator for internal migration (Fan 1997; Xing, Fan et al. 2009; Chen and Groenewold 2011; Howell 2011). For instance, rural and urban level of regional inequality has played an important part in the massive internal migration (Li, Liu et al. 2014; Liu, Stillwell et al. 2014). This can be reflected by the large share of rural-urban migrants in the total internal migration, which has been happening for the past few decades.

Many argue that China's regional inequality results from the unbalanced regional development policies, market forces, and (human and natural) resource endowment differences (Tian 2001; Zhang and Song 2003; Wang and Piesse 2010; Sheng 2011). For instance, Wang et al. (2013) argue that region-biased policies favouring the development of cities have led to the under-provision of social public goods in the countryside, which could be one trigger for rural-urban migration. This is in line with the new economics of labour migration, wherein regional inequality could stimulate the mobilization and redistribution of production (Clark 1982; King 2012). As an important element of production, labour or human capital (including the knowledge and entrepreneurship stocked in the labour force) also obeys economics laws, moving from areas of lower to higher income across the country in pursuit of better life opportunities (Taylor, Arango et al. 1996; De Haas 2007b; Jenicek 2010; Ye, Wang et al. 2013; Liang, Li et al. 2014; Liu, Stillwell et al. 2014; Liu, Qi et al. 2015).

This sub-section has shown the effects of the regional development policies, and their associations with the internal migration in China. These brief descriptions of regional development policies are aimed at providing the background for the unbalanced regional development, which sets the scene for the interprovincial migration to take place. Based on this, the next sub-section will utilise the learning achieved so far in this chapter to describe the recent changes of China's internal migration in order to better situate the research questions of this thesis.

2.3.1.4 Recent changes of China's internal migration

Despite the long-standing two main characteristics of internal migration in China[24] – namely, being temporary in nature and being strictly regulated by policies – China's internal migration is experiencing important structural changes and thus new trends are emerging. According to Reports on China's

[24] See Sub-section 2.3.1.1 for more details.

Migrant Population Development and China Statistical Yearbook, China's internal migration has shown seven major recent changes since 2010, which will be explained in detail in below.

First, even though the total volume of internal migration remains vast, it has seen a slow decline since 2015[25] as a result of a continued decelerating natural population growth that is marked by slackening fertility and declining mortality. However, internal migration continues to be the leading factor in shaping population changes in China, and is likely to remain so for a long time in the future. If put into Zelinsky's mobility transition theoretical framework (Zelinsky 1971), this new development signposts the onset of 'Phase C - The Late Transitional Society' in 'The Vital Transition' or 'Phase II - The Late Transitional Society' in 'The Vital Transition' for the Chinese society (Zhu 2018). Most importantly, this new development may have marked a significant structural and compositional change in this late transitional phase of Chinese society: the continued decades' growth of rural-urban migration has finally levelled off, whilst urban-urban migration is gaining momentum and becoming an increasingly important component in internal migration of China (Zhu, Lin et al. 2016; Zhu 2018). This may set China's internal migration apart from its peers in many modernising societies such as India and Vietnam (Nath 1988; Chen and Ravallion 2004; Deshingkar 2006; Phuong and McPeak 2010; Anh, Rigg et al. 2012; Czaika 2012; Nguyen and Locke 2014), as important broad patterns characterising internal migration within these developing countries are exactly the increasing predominance of rural-urban migration and large volumes of marriage-oriented rural-rural migration (Lucas 1997). For instance, there has been a huge increase in internal migration (primarily rural-urban migration) in India in the last decade, whose total number raised from 228 million in 1993 to 352 million in 2007-2008 (Parida and Raman 2018). Therefore, the current internal migration in China offers a valuable opportunity to advance internal migration research and to understand the structure and component of the migration

[25] According to China Statistical Yearbook 2017, the peak value was 298 million in 2014, which reduced to 294 and 292 million in 2015 and 2016 respectively.

system at the critical time point of mobility transition. It is also now most obvious to analyse the important role of urban-urban migration in this transitional process.

Secondly, there are signs of increasing stability, as the average settlement period for migrants grows from 4.8 years in 2014 to 5.7 years in 2016[26]. Meanwhile, the proportion of intraprovincial migration has started to increase slowly, although interprovincial movements still maintain a dominant share for rural migrants between 2010 and 2016. This new development not only reflects that China's internal migration has entered the stabilising stage where migrants are increasingly seeking for settlement, but also has positive impact upon data-driven empirical migration research, as temporary and circular migration has been most prevalent in modernising societies and has led to poor statistical base for migration studies in general (Lucas 1997). Moreover, this new development may be relevant to the fact that Chinese economy keeps growing as capitalist development proceeds and regions become more integrated with widespread grass-root private entrepreneurships alongside state-owned entities gaining momentum (Yiu, Lau et al. 2007; Fielding 2010). Consequently, Chinese economy is becoming more capable of retaining 'fluid' labour on the local level overall, when it also starts to enjoy economies of scale as a whole (Appadurai 1990; Fukuyama 1995).

Thirdly, there has been continued growth of new-generation migrants (born in 1980 and after), whose share has risen from 48.8% in 2013 to 64.7% in 2016[26]. Compared with migrants of the former generation, this new-generation migrants are not only younger, but also tend to have little agricultural experience and thus have a lesser sense of belonging to rural identity (Wang 2001). Many of the new-generation migrants, in particular, prefer to relocate to bigger cities in seek of employment opportunities with better income prospects but more importantly with better career trajectories (Wang 2001; Huang and Zhan 2005). Therefore, this new development is also unsurprisingly accompanied by increases in education and vocational training among migrants, with evidence from the trends in, and compositions of, the interprovincial flows to and from big cities such as Beijing and Shanghai (Huang and Zhan 2005;

[26] Data source is Reports on China's Migrant Population Development.

Fielding 2010). In this respect, since the reform in 1978 China's internal migration has shown some different features than that of India, which has also seen large-scale internal migration since early 1990s. The internal migration in India, however, is mainly driven by poverty, hunger and unemployment, as most of the internal migrants are low-skilled or unskilled and conducting moves is therefore an important approach to search for survival (Parida and Raman 2018).

Fourth, internal migrants have experienced a faster aging process than the national total population. Specifically, the average age of migrants has grown from 27.3 to 29.8 years old from 2011 to 2016[26], whilst the national total population aged about 2.0 years during this period. This fast-aging process of the total population in China is widely recognised as one consequence of strong state intervention with population policies such as the notorious one-child policy (Chen and Powell 2012). China is indeed getting old before it gets rich, as an aging population is typically seen as a burden rather than a blessing to socio-economic development (Fan 2007; Chen and Powell 2012). Therefore, this fast-aging migration population could inevitably cause systemic and structural changes with reference to the internal migration system, such as household migration decision-making and the management and design of social support system in both the origin and the destination (Chan and Zhang 1999; Rozelle, Taylor et al. 1999; Fan 2007). This demographic change of the migration population may also relate to the aforementioned observable signs of increasing stability of the internal migration system - as migrants get older, they tend to have stronger intention and desire to settle down (either at their chosen destinations or go back to their origins) to start a family or return to care responsibilities for family members (Rogers and Castro 1981; Rogers and Willekens 1986; Wang and Fan 2006; Schanbacher 2007).

Fifth, internal migrants have enjoyed a bigger household income growth than the national average. To be precise, there has been a 72.6% increase between 2013 and 2016 for migrants[26], in contrast to the national average income growth rate of 30.1% during the same period[27]. This is in line with the

[27] Data source is the National Bureau of Statistics of China.

aforementioned neoclassical economic paradigm, which expects population movement originates from low-income places and settles in (comparatively) high-income destinations with migrants seeking to maximise their utility in this process (Sjaastad 1962; Massey, Arango et al. 1993; King 2012). Higher economic returns are indeed the fundamental booster powering China's long-standing and massive nationwide population relocation process (Fan 2005b; Fan 2007). Such higher economic returns also have important impacts upon households such as increasing family income through remittance, empowering female migrants and restructuring family relations (Fan 2003; Wang and Fan 2006; Fan and Wang 2008; Fan, Sun et al. 2011; Wang and Fan 2012). Migration economic returns may also create spill-over effect of consumption and entrepreneurship in the local neighbourhood and community of the origins (Rozelle, Taylor et al. 1999; Wang and Fan 2006; Fan 2007; Snyder and Chern 2009; Gao 2012; Wahba and Zenou 2012; Han and Chen 2016; Cao, Li et al. 2017). Nevertheless, none of these would have happened without the onset of the famous 'Reform and Open-up' policy in China since 1978, which opened the Chinese market to international capitals (Prahalad and Lieberthal 1998; Fujita and Hu 2001; Nee, Opper et al. 2007; Tsai 2007).

Sixth, there has been a growing trend of familisation among internal migrants with rising household size from 2.5 to 2.6 persons between 2013 and 2015[6]. This new development of China's internal migration stands as a contrast to the prevailing conjugal separations and family splits observed in China among earlier migrants for three decades (Cheng 1991; Ma and Fan 1994; Fan 1995; Gordon 1995; Wang and Fan 2006; Fan and Sun 2008; Fan and Wang 2008; Fan, Sun et al. 2011; Ye, Wang et al. 2013). This growing trend of familisation among internal migrants therefore lends further support to the hypothesis that China is currently in the late transitional stage based on Zelinsky's mobility transition theoretical framework (Zelinsky 1971; Zhu 2018). There are mainly three groups of left-behind family members when the migration decision is made within the household, namely wives, the elderly and children (Biao 2007; Fan 2007; Wen and Lin 2012; Ye, Wang et al. 2013). The left-behind population encounter various problems, such as the management of agricultural production, support and care for the elderly, and children's education and well-being (Chang, Dong et al. 2011; Wen and Lin

2012; Ye, Wang et al. 2013; Lu, Lin et al. 2016). Compared with non-migrant households, however, the situation of the left-behind population of migrant households in general is not much worse (Biao 2007), which may be relevant to the remittance sent back by the migrants to the household (Deshingkar 2006; Chang, Dong et al. 2011; Cong and Silverstein 2011). It is widely believed that the fundamental cause of conjugal separations and family splits is the long-standing Hukou system, state urban-preference and regional unbalanced development policies, although on the surface family separations within migrant households may seem like voluntary decisions at the individual level (Fan 1997; Biao 2007; Fan 2007). Apart from demographic factors such as age, gender, health status and education (Rogers and Castro 1981; Rogers and Willekens 1986), these institutional factors set fundamental constraints on household migration decisions of who migrates and who stays back (Biao 2007). Therefore, this new development of migration familisation may be relevant to the increasing stability of China's internal migration, as family migration is generally seen as a positive prelude sign to permanent settlement. There are signs that familisation of migration has contributed to reducing the gender preference of males in the migration process within China (Fan and Huang 1998; Fan 2003; Raghuram 2004), whilst more female internal migrants are observed in the 1970s of Latin America (Lucas 1997). Family migration may also be a possible solution to reduce the hardship encountered by the left-behind population within migrant-households (Wen and Lin 2012; Ye, Wang et al. 2013).

Seventh, the coverage of social welfare has seen observable improvement for internal migrants. For instance, 91.1% of internal migrants have joined at least one type of health insurances in 2016, standing as a striking with the less than 30% coverage rate in 2011[26]. This is in line with the overall enhancement of social welfare coverage across China in recent years (Saich 2000; Liu 2018), which is also a part of the national development strategy to reduce regional inequalities (Zhang and Kanbur 2005; Fan and Sun 2008). It is definitely a positive development for migrants, which reduces their living expense and pressure and probably has helped them to achieve familisation and stability in the process of relocation (Keung Wong, Li et al. 2007; Wang and Fan 2012).

On the whole, these seven major recent trends of China's internal migration not only show important development of the migration process, but also migrants' changing demographic features and the impacts of migration. As with the changes of the migration process, the recent development has signposted the onset of the late transitional stage with increasing stability and familisation of the migrants, along with enhanced social welfare support. These new trends reflects China's socio-economic development, which may mark an ongoing critical transition from underdevelopment to development for China (Kissinger and Hormann 2011; Shambaugh 2013). Moreover, the internal migration population in China has a growing share of the new-generation but still shows signs of faster-than-national-average aging and other demographic features, which is closely related to the recent industrialisation process and socio-economic changes in China. Such demographic changes of migrants pose new opportunities and challenges for the Chinese government in managing urbanisation and developing the social welfare system (Davin 1996; Sjoberg 1999; Feng 2002; Biao 2007).

Most importantly, the internal migration population has experienced a markedly large income growth than that of the national average, which has long-lasting implications for further economic growth based on arguments of the neoclassical economic paradigm (Todaro 1980). Similar to most developing countries (Lucas 1997; Parida and Raman 2018), the importance of internal migration in China is indeed not necessarily in its process or effect, but actually in its implications for economic growth (Todaro 1980). In this respect, studying internal migration of the largest developing country, namely China's internal migration, could provide valuable lessons for other developing countries, whilst researching on translating internal migration into a factor contributing to Third World underdevelopment remains as a prevailing and important task (Zhu 2018).

2.3.2 Drivers of internal migration in China: why is the economic perspective important?

The last sub-section has elaborated the characteristics and major recent trends of China's internal migration. This sub-section endeavours to explicitly explain both macro- and micro-level drivers of internal migration (Zelinsky 1971; Wood 1982; King 2012) in China.

2.3.2.1 Macro-level Drivers of China's internal migration: rural and urban divide, and regional inequality

The explanation of China's internal migration macro-level drivers in this sub-section draws on the formerly learnt knowledge of the three perspectives of migration theoretical analytical frameworks - economic, socio-cultural and neo-Marxism. Among macro-level drivers of China's internal migration, some may have played a predominant role and have been most effective in relocating the migrating population across different regions, such as economic factors and regional development policies (Fan 1996; Fan 1997; Fan 2005a; Shen 2012); by contrast, other drivers could be peripheral in certain areas or regions at a certain time and could have varying impacts upon different age groups of the population or some particular households (Ye, Wang et al. 2013). For instance, Wu and Yao (2003) showed that industrial structure change and the non-agricultural industry growth in rural areas had significantly affected migration decision-making among Chinese peasants between 1995 and 1998. This sub-section therefore serves as an extensive discussion of different macro-level migration drivers of China's internal migration, so as to further clarify why this thesis chooses migration drivers of the economic perspective as the major focus.

Following the economic reform in 1978, large scales of internal migration have started to emerge and spread all over China. The economic reform therefore has been commonly regarded as symbolising the advent of initial large-scale internal migration in China's recent history. Theories of the neo-Marxism approaches have been most useful in explaining the onset of the recent massive internal migration in China after the economic reform (Fan 1995; Chan 2013; Ye, Wang et al. 2013), which attributes to the joint forces of some most primitive macro-level drivers such as the implementation of 'Reform and Open-up' policy and the subsequent penetration of international corporate capitalism (Prahalad and Lieberthal 1998; Fujita and Hu 2001; Nee, Opper et al. 2007; Tsai 2007).

Regarding the continuously growing mobility rate in China over the past four decades, it is of particular importance to discuss regional inequality as a major driver in order to better understand how mechanisms, patterns and dynamics of interprovincial migration play out. By drawing lessons from prior studies and incorporating labour migration theories with China's reality, this thesis argues that regional inequality is a major macro-level driver of China's internal migration. Such an argument is also in line with the established literature, wherein regional inequality has been put forward as a major driver of interprovincial migration in China (Fan 2005a; Shen 2016b). In empirical studies of international settings, for instance, a strong association between regional inequality and migration flows has been repeatedly identified (Vignoli 2008; De Haas 2010b; Ye, Wang et al. 2013). The predominant role of the regional inequality has been stressed not only in different settings and contexts (Ye, Wang et al. 2013), but also by a wide range of migration theories - Lewis's unlimited supply of labour (Lewis 1954b), Harris and Todaro's expected income (Harris and Todaro 1970) and Stark's relative deprivation (Stark and Taylor 1991).

China has experienced severe regional inequality for many years (Jian, Sachs et al. 1996; Fan and Sun 2008; Fan, Kanbur et al. 2009; Wang, Piesse et al. 2013), which have been well reflected by the scale and direction of internal migration (Fan and Sun 2008; Ye, Wang et al. 2013). Regional inequality therefore has been widely regarded as an effective predicator for the direction and volume of internal migration flows in China (Fan 1997; Xing, Fan et al. 2009; Chen and Groenewold 2011; Howell 2011). Regional inequality in China is recognised as to have three major dimensions – rural and urban, province, and region (Fan and Sun 2008). All these three dimensions of regional inequality are believed to be closely related to the unbalanced regional development policies[28] adopted in China (Tian 2001; Zhang

[28] See sub-section 2.3.1.2 and 2.3.1.3 for more explanation.

and Song 2003; Wang and Piesse 2010; Sheng 2011). For instance, the rural-urban inequality of the provision of social public goods has clearly stemmed from urban-oriented regional development policies (Wang, Piesse et al. 2013). Another example is the existence of ‘escalator regions’ in China, which are more socio-economically developed than other regions and have particularly strong positive relationship of social and geographical mobility in comparison (Fielding 2010).

Regional inequality could also influence internal migration through conditioning macro-level demographic characteristics such as education, which also has an important impact upon migrants’ accessibility to employment opportunities and economic returns of human capital on the aggregated level. For instance, Yu, Xu et al. (2014) revealed that the economic return for migration varied greatly across different geographic regions, verifying that more developed regions tended to have more employment opportunities and higher returns to human capital and weaker institutional and cultural barriers for migrants. Such a finding is also related to the important concept of the ‘escalator region’[20], which is featured with strong both in- and out-migration flows but with distinctive age structures and socio-economic statuses (Fielding 1992). Indeed, there is strong evidence for the existence of a positive relationship between social and geographical mobility in mainland China. Based on the 2000 and 2005 census data, interprovincial migrants who moved to ‘escalator regions’ such as Beijing and Shanghai were particularly more likely to search for occupational promotion and improved social status (Fielding 2010). Other regularities between education and internal migration on the macro-level observed around the world also hold true in the case of China’s internal migration. For instance, Li (2009) found that interprovincial migration and education had a relationship of inverted-U shape, with the number of interprovincial migrants peaking at middle-school (including other equivalent) education; such a relationship was particularly pronounced among rural-urban migrants. Liu, Stillwell et al. (2014) found that young people were more likely to migrate in response to economic inequalities, and that educated and skilled migrants showed more signs of clustering in more socio-economically developed destinations than less-educated ones.

Whilst each research chapter (4, 5 and 6) will specifically be designated to discuss the association of migration and regional inequality on each of three major dimensions (rural and urban, province, and region) respectively, some other political structures other than the unbalanced regional development policies could nevertheless serve as substantial migration drivers or barriers and will be discussed here. Some of these macro-level political structures could penetrate into household or individual migration processes. For instance, Yan, Bauer et al. (2014) found that China's unique land ownership and tenure system had played a significant role in farmers' migration decision-making process within the household. Other macro-level political structures, such as the Hukou system, could have complicated effects in shaping and conditioning internal migration flows (Chan 2010b; Bosker, Brakman et al. 2012). As explained in sub-section 2.3.1.1, China's internal migration has two major characteristics of being temporary in nature and being regulated by heavy institutional interventions, which have very close linkages to macro-level migration drivers such as regional development and migration management policies. As regional inequality as a major macro-level migration driver has been extensively discussed, it is therefore important to address the complex roles of the Hukou system in China towards the management of internal migration in the following. This is also due to the Hukou system's strong influences upon internal migration in terms of directions, volumes and durations.

The complicated effects of the Hukou system have led to mixed explanations and competing discourses among migration policy studies. Part of the reason is the very dynamic and always-changing migration management policies in China. Taking the Hukou system and the regional development policies for example, they have been most dynamic and varied greatly across different regions[29] (Chan 2009). Overall, there has been continued heavy reliance upon institutional interventions such as the Hukou system in China, due to the distrust in the role of marketerisation in relocating and redistributing population (Chan 2009; Wang and Fan 2012; Liu, Stillwell et al. 2014). To illustrate, Wang and Fan (2012) showed that the Hukou system still posed as a deterrent for migrant workers, despite the recent

[29] Sub-section 2.3.1.2 discusses the origins and recent changes of the Hukou system and the regional development policies in more detail, whilst this sub-section focuses on broadly explaining their mixed roles in migration management and how they can act as impeding deterring factors for internal migrants.

relaxing of its constraints upon the settlement of migrants. Liu, Stillwell et al. (2014) therefore argued that the state was not successful in either alleviating regional inequalities or achieving migration system equilibrium through such institutional interventions, regardless of its recent great efforts in constantly re-adjusting and reforming these political structures.

Contending with the wide belief of the deterring effect of the Hukou system upon internal migrants, some scholars argue that the combined effects of both national and local migration management policies have led to mixed but overall positive outcomes for the socio-economic development of the country as a whole (Zhan 2011; Naybor 2018). However, they all agree that Chinese planners and policy makers should acknowledge the value and contribution of internal migration and provide more institutional recognition and support to migrants (Zhan 2011; Naybor 2018). In other words, more migration policies should be localised and designated tailoring for the needs of migrants within the regional context. One major reason to support this argument lies in the belief that the current Hukou system causes little deterring impact upon internal migrants (Zhan 2011; Naybor 2018). To be more precise, Zhan (2011) argued that the Hukou system had fallen in significance and abolishing it could help little in improving the life quality of migrant workers. Similarly, one recent study shows that there are few rural-to-urban migrants who are willing to give up their rural Hukou in order to obtain the urban Hukou, although the Hukou policies are still changing and there is a lack of data on explaining the impacts of the latest changes (Naybor 2018).

Scholars from both sides, however, agree on the continued adjustment and reform of migration management policies. One solution is to speed up the citizenisation process of all society members. For instance, Wu and Xiao (2014) confirmed that the citizenisation of rural migrant workers can expand employment and the scale of investment, accelerate structural adjustment, and boost economic growth. Another approach lies in addressing the long-standing concern about the link running from unmanaged migration to rising unemployment in the destination. Liu (2012) proposed that rural-urban migration should not be held responsible for urban unemployment in China. Althought some job opportunities may be taken away from local urban residents and the equilibrium local wage might be drawn down by rural-urban migrants, these rural-urban migrants could help to expand the production and consumption capacity and increase the total local economic outputs. Positive impacts of rural-urban migrants could therefore outweigh negative ones overall for the urban destinations. In this sense, encouraging and fostering the pair-wise connection between the origin and the destination could be enormously beneficial with respect to managing rural-urban migration. Taking the Hukou system as an example, Wan (1995) suggested that the facilitation of two-way rather than one-directional flows should be implemented.

2.3.2.2 Micro-level Drivers of China's internal migration

This sub-section will explain how micro-level migration drivers play out in China's internal migration. On the whole, findings on micro-level migration drivers of China's internal migration confirm the those of existing empirical studies on internal migration in other countries. The following will provide evidence for it by extensively reviewing relevant empirical studies of micro-level migration drivers in China, including micro-level demographic factors and individual's migration experience.

A large body of studies have elaborated upon demographic characteristics of migrants across and above different stages of migration decision-making process [30] for China's internal migration, such as population structure, employment, identity, physical and mental health, social welfare and consumption. Regarding demographic characteristics of individuals, impacts from aspiration, education, skill, gender, age, past migration experience and other characteristics of migrants are rather evident (Cheng 1991; Yang 2000; Liu 2005; Poncet 2006; Chen, Liu et al. 2013; Piotrowski and Tong 2013; Ye, Wang et al. 2013). For example, Liu and Shen (2014a) found that Chinese skilled workers valued career prospects more over life qualities when making migration decisions. Qu and Zhao (2014) argued that rural

[30] It refers to: Stage 1, when migrants first decide to leave their origin; Stage 2, when migrants start to search for destinations; and Stage 3, when migrants finally choose among all the alternative destinations (Rossi and Shlay 1982). See sub-section 2.2.2.2 for more explanation.

migrants depended a lot upon former migration experiences in the migration decision-making process. Chiang, Hannum et al. (2014) pointed out that migration decisions of young migrants were profoundly influenced by gender as boy and girl migrants behaved substantially differently in the process. Similar to internal migration in the world, age selection is also evident in the case of internal migration in China. For instance, the study of Wan (1995) revealed that family migration was significantly negatively influenced by age, that marriage had a much greater deterrent impact on individual migration, and that family network was a more significant determinant for individual rather than family migration. This also points to the strong empirical evidence argued by Yang and Guo (1999) that individuals often do not act alone and that migration could be a group decision within the household and community in China. Such a family strategy of migration of Chinese internal migrants links closely with observations around the world, regarding internal migrants' life course, family norm and social expectations in terms of fertility, education, marriage, inheritance, and risk spreading (Findley 1977).

Apart from the age seclectivity of migration, the gender preference of males is also prevalent in China, whilst male and female migrants tend to behave differently in the migration process. For instance, Yang (2000) revealed that males were more inclined to conduct individual migration than females. Wan (1995) found that females took extremely inactive part in migration, and that individual females also tended to be sensitive to migraotry distance and moved for short distances if they conducted any migration at all. This is in line with the general finding in migration research literature that males tend to have a prodominant role in migration, especially for pioneering or long-distance migration (Findley 1977), whilst a large volume of female migrants take part in marriage-oriented rural-rural migration with short migratory distances (Lucas 1997; Fan and Huang 1998). Gender could also co-function with other factors to influence migration activities profoundly. For example, there exists strong evidence showing that household structure exerts more influences on female rather than male migration by limiting female migrants' participation in economic activities (Fan and Huang 1998; Yang and Guo 1999; Fan and Li 2002; Kofman 2004; King and Skeldon 2010). Such a gender selection of migration with a male preference is also common in many other countries (Lucas 1978). Conversely, however, Zuo (2008) argued that within the household joint decision-making was favoured by male adult out-migrants after marriage, and that married women usually had more say in the migration decision-making as they tended to take more household management responsibilities. These research findings of Zuo (2008) contradicted with other studies where females were found to be depowered one way or another when making migration decisions (Liu 2012; Zhang, Gao et al. 2013; Zhang 2013a; Chiang, Hannum et al. 2014; Nguyen and Locke 2014).

In terms of individual's internal migration experience, similar patterns are observed in China's internal migration. Internal migrants not only suffer from aggregated-level exclusions from both local society and institutional systems in the destination as a social group, but also encounter within-group descrimination from peer migrants on the individual level. For instance, Zhan (2011) showed that not only the local residents and job markets but also institutional systems such as the Hukou and social welfare system widely held social discrimination against migrants, and that social exclusion and discrimination even existed among different migrant ethnic groups and among migrants of the same ethnic group themselves (Wang and Fan 2012; Lu and Zhou 2013; Shin, Wan et al. 2013; Ling 2015; Wang, Guo et al. 2015a, 2015b). As socio-economic, institutional and cultural barriers persistently exist to migrant workers and greatly affect their migration behaviour and life quality, it is therefore common to find migrants suffering from physical and mental health such as occupational injuries and identity crisis. For instance, many studies have confirmed that migrant workers shoulder a disproportionately heavy burden of occupational injuries and mortalities as well as higher risks of other non-occupational diseases in China (Fitzgerald, Chen et al. 2013; Sun and Liu 2014; Yang, He et al. 2015; Kang, Xiao et al. 2016; Wang, Hu et al. 2016). Gui, Berry et al. (2012) showed that similar to international immigrants, integration was also the best strategy for seasonal migrant workers to acculturate and attain social wellbeing. Nevertheless, mental health is increasingly a problem (Wong, He et al. 2008; Chan and Qiu 2011; Qiu, Caine et al. 2011; Zhu, Geng et al. 2013; Zhong, Liu et al. 2015), without the sufficient support of medical services for migrants in the host society (Zhuang 2009; Dai 2011; Qiu, Caine et al. 2011; Xu, Guan et al. 2011; Ngok 2012; Sun and Liu 2014). Indeed, individual migrants differ in demographic and socio-economic characteristics and hold different expectations of residence places, which determines the satisfaction degree of the conditions in the current destination for migrants and their awareness and strategy to assimilate to the destination society. All of these attribute to their different migratory behaviours.

In summary, micro-level drivers of China's internal migration show consistencies and similarities to those of internal migration in other regions and countries. Most importantly, two major characteristics of China's internal migration (temporariness and heavy institutional regulation) have very close linkages to macro-level migration drivers such as regional development and migration management policies. These macro-level drivers of China's internal migration are typically external to either individual migrants or their households or local communities, and are very unique to China. This thesis then decides to focus on analysing the role of macro-level factors (more specifically, the regional inequality) in understanding the unique processes of China's internal migration. In view of this, the next sub-section will explain explicitly why the economic perspective of internal migration drivers is important for China.

2.3.2.3 Drivers of China's internal migration in China: why is the economic perspective important?

As China is currently going through a structural mobility transition from perpetuation to stabilisation[31], it provides an excellent opportunity to further our understanding of the development trajectory of migration on the aggregated level (Zhu 2018). This thesis therefore takes up this opportunity and challenge and endeavours to improve the knowledge of how such a mobility transition unfolds in a socialist transitional economy by primarily focusing on macro-level economic factors and regional differences (Fan 1999). This is because the macro-level economic factors along with regional development and migration management policies have the most profound impact on China's internal migration overall (Barro and Sala-i-Martin 1990; Barro and Sala-i-Martin 1992a, 1992b; Cai and Wang

[31] See Sub-section 2.2.1.2 for more details.

2003a, 2003b; Bao, Shi et al. 2005; Song and Wang 2005; Wang, Wei et al. 2005; Li 2009; Shen 2012; Shen 2013). These macro-level factors are not only crucial to the initialisation of internal migration in China but also remain influential in the perpetuation stage of migration, and continue to be the decisive force in transitioning the current mobility structure into the stage of stabilisation. Yan (2007) provided strong evidence for the importance of these macro-level factors to the development of internal migration in China, who established a ‘flow chain’ migration model based on the population mobility theories and pointed out that regional economic inequality was the initial stimulus whilst the net expected income minus the migration cost to the potential destination was actually the determining factor for China’s interprovincial migration. Moreover, Yan (2007) emphasised that the recent regional economic inequality originated from the unbalanced implementation of ‘Reform and Open-up’ policy across different regions and the consequently uneven distribution of wealth. The macro-level economic factors along with regional development and migration management policies therefore consist of the fundamental context in studying the interprovincial migration of China. One underlying assumption of this thesis is that China is likely to arrive at the stabilisation stage of migration[32] soon based on the evidence of recent migration trends[33], as this structural transition of mobility is important and interesting to research on; unless China is going to make some severe mistakes in the near future to stop the current socio-economic development trend, such as crushing productivity with excessive state control, embargoing international trades entirely, or initiating and engaging in large-scale military confrontations and wars. Under these rare conditions, however, the stabilisation of mobility and migration in the general population might not take place in the internal migration system of China due to such grave socio-economic interruptions and crises.

The implications of the findings presented in this whole sub-section of 2.3.2 is that economic and institutional factors are major migration determinants that play significant parts in internal migration

[32] It means that China will have a stable mobility rate of internal migration with organised and ordered urban hierarchies upholding the distribution equilibrium of the factors of production (Zelinsky 1971; Wood 1982; Skeldon 2014). Other key demographic features include continuing low fertility and mortality rates with very little population increase, and internal migration will be the primary source of population change (De Haas 2007a).
[33] See sub-section 2.3.1.4 for more details.

processes of China. To elaborate, China's internal flows are spatially imbalanced primarily due to macro-level regional inequalities and institutional policies, whilst these internal flows are powered by individual migrants substantially differing in migratory behaviours and experiences. The ultimate goal of any migrant management policies should always be made to be fit to the needs of the local society and compatiable with economic development. It is therefore vital that China must learn from other countries' experience in redistributing population with a much greater degree of efficiency. As the stabilisation of mobility with improved movement freedom is likely to be the foreseeable future for China, it is crucial to understand how to harness the positive impacts of migration whilst minimising its negative side-effects.

2.4 Discussion

By reviewing characteristics, patterns and trends, as well as major drivers of internal migration, this chapter has discussed how the evolving theories and empirical knowledge in labour migration are providing new opportunities for studying interprovincial migration in China. Most importantly these theories and methods have pointed to three promising research directions: the first one is to investigate the rural and urban components of provincial income, which is also closely related to multi-directional movements between provinces (i.e. urban-urban, urban-rural, rural-urban and rural-rural interprovincial flows); the second one is to quantify flow dependencies; and the third one is to explore the role of distance in the migration process by considering the contiguity of provinces and regional differences.

The first of these will be carried out in Chapter 4 by examining the four interprovincial migration flow types through rural and urban income components, in order to answer the first research question 'What are the associations between China's interprovincial migration and rural and urban level of regional inequality through origin and destination population, income and distance?'. The second research direction is to be examined in Chapter 5, in order to answer the second research question 'What are the associations between China's interprovincial migration and province level of regional inequality through flow dependencies?'. Chapter 6 will extend the specification of distance decay to the non-linear function along with considering the contiguity of provinces and regional differences, in answering the third research question 'What are the associations between China's interprovincial migration and region level of regional inequality through distance decay?'.

This chapter has shown how it is now much easier to study interprovincial migration in China, by enabling this thesis to be better embedded within the development of theories and methods of labour migration. This thesis aims to contribute to bridging the three knowledge gaps as identified in this chapter, namely the different interprovincial migration flow types (Chapter 4), the flow dependencies (Chapter 5), and the non-linear rate of distance decay that are subject to the contiguity of provinces and regions (Chapter 6). All the evidence in this chapter points towards a clear research plan to tackle these three under-researched issues surrounding the labour migration in China. With these newly identified opportunities in mind, the following chapters will take up the challenges that they entail in order to achieve a better understanding of China's interprovincial migration.

Chapter 3 Data and methodology

This Chapter starts by explaining how the appropriate data are compiled from multiple sources to develop new measurements addressing the three research questions[34]. It then moves on to describe the evolvement of the gravity model of migration and write out the general form of linear regression. The final section summarises the learning of this chapter, setting the scene for model extensions in the following results chapters.

3.1 Data and measurements

The core research interest of this thesis is to investigate what the associations between China's interprovincial migration and three levels of regional inequality (rural and urban, province and region) are. It is important to ask and answer this question in order to have a richer and more systematic understanding of China's internal migration. This in turn will help politicians to make better informed policies to reduce regional inequality and improve migration governance.

The investigation of the associations between China's interprovincial migration and three levels of regional inequality is conducted based on extensions of gravity models (Christian and Braden 1966; Claeson 1969). Some key predictors are measurements of the masses of origins and destinations (measurements which include population and income) and the distance between the two locations. This thesis will follow principles of traditional gravity models and adopt these key predictors in designing models. This section details the overall development of measurements for key predictors such as population, income and distance.

[34] See Section 1.2 of Chapter 1 for more details.

3.1.1 Migrants and migration flows

Chinese censuses usually have two datasets – short-form and long-form, with short-form covering the whole population whilst long-form taking a sample of the whole population. In the 2000 Census, for instance, the long-form is a 9.5% sample of the total population and migrants are recorded differently in the short-form and long-form. Unlike earlier censuses, both 2000 and 2010 Census enumerate the number of migrants at their usual/actual place of residence rather than their original registered Hukou address (Fielding 2010). To be more specific, interprovincial migrants are recorded in the long-form by eight different lengths of stay in the destination provinces – more than six months but less than a year, one year, two years, three years, four years, five years, more than five years but less than lifetime, and lifetime. The short-form of the 2000 Census, by contrast, provides only the total counts and does not distinguish interprovincial migrants by different lengths of stay (as long as they stayed in the destination provinces for more than six months). In the 2010 Census, the long-form is a 10% sample of the short-form that covers the whole population. However, migrants are also recorded differently in the short-form and long-form of the 2010 Census. In the long-form of the 2010 Census, interprovincial migrants are recorded by two different lengths of stay in the destination provinces – more than five years but less than lifetime, and lifetime. In the short-form of the 2010 Census, interprovincial migrants are recorded only as the total counts (the total number of migrants who have stayed in the destination provinces between six months and lifetime) and are not distinguishable by different lengths of stay in the destination provinces.

In order to adopt a consistent measurement of interprovincial migration in 2000 and 2010, the total migration data are taken from the short-form in the Census for both years. I am also interested in studying the four interprovincial migration flow types, but such data are only available in the long-form dataset of the 2010 Census. Therefore, the data for the four interprovincial migration flow types are taken from the long-form dataset of the 2010 Census, which are also lifetime (but more than six months) migration counts and provide the data needed to perform the analyses that follow. Consequently, an interprovincial migration flow in this thesis refers to the number of migrants who have left an original Hukou place (their originally registered legal residence) and stayed at a destination of a different province for at least six months. However, an individual that has moved to another county for more than six months will not be recognised as a migrant once this individual obtains the local Hukou. This in theory could lead to the issue of endogeneity as individuals may delay updating their household registration information to the authorities. But it is not a real problem in census data collection, as the coverage rate is almost 100% in both censuses (98.19% in 2000 and 99.88% in 2010) and the collection of the household registration information is dictated by the census data collector rather than individuals[35]. Under this definition, an interprovincial migration flow bears some resemblance to a migration stock, which also accounts the number of migrants based on the difference between birthplace and residence place (Bell, Blake et al. 2002). But interprovincial migration flow and stock are two distinctive concepts in fact. In the context of this thesis, for instance, an interprovincial migration flow originating from Province A and ending at Province B contains lifetime (but more than six-month) migrants, whilst an interprovincial migration stock of Province A is equal to the sum of all the interprovincial flows ending at Province A.

One limitation of adopting such a definition lies in its inability to count multiple and return moves (Bell, Blake et al. 2002). This is because the time period in migration is essential to defining and categorising different migration processes. The prevailing focus in the academic community currently has been a distinctive temporary/permanent or movers/stayers classification (Meeus 2012; Pine 2014; Robertson 2014; Aybek 2015), whilst an extended three-category classification that emphases and attempts to distinguish one-time movers, multiple movers, and stayers from each other has been rising in popularity recently (Silvestre and Reher 2014). Some scholars even argue that migration is undergoing some fundamental change in its nature around the global, with rising salience in being more circular and temporary (Widdis 1988; Castles 2002; May and Thrift 2003; Meeus 2012; Griffiths, Rogers et al. 2013).

[35] Individuals are required to show their household registration certificates to the census data collector in China, in order to ensure the identities of the individual and the accuracy and reliability of census data collection.

Regarding the category name of multiple movers in particular, some scholars prefer to call them return, circular or repeat migration (Wang and Fan 2006; Hunter 2011; Agadjanian, Gorina et al. 2014; Jia and Liu 2014). These different classifications could get more diverse when attempting to depict complex migration realities (Khoo, Hugo et al. 2008; Meeus 2012). To be more specific, in the scientific community there exists continuing attention and interest to its conceptual significance upon flows (Yapa and Wolpert 1971; Rogerson, Raymer et al. 1990; Smith, Raymer et al. 2010), ruptures (Widdis 1988; Rogerson, Raymer et al. 1990; Rogers, Raymer et al. 2003; Robertson 2014), cycles and synchronicity (Newbold 2005), life course (Khoo, Hugo et al. 2008; Meeus 2012; Pine 2014; Robertson 2014; Silvestre and Reher 2014; Aybek 2015), as well as other longitudinal patterns (Yapa and Wolpert 1971; Schmeidl 1997; Blacklock, Heneghan et al. 2012; Lomax, Stillwell et al. 2014). Most importantly, different treatments of time periods could alter the results of migration empirical studies. As Molho (2013) stressed, temporal factors, such as the length of staying and life cycle effects, are essential in analysing migration. This is particularly true for defining migration flows. For example, the number of migrants could be remarkably different when data is collected over different time intervals. For instance, Rogers, Raymer et al. (2003) found that widths of selected time intervals have a significant impact upon the number and portion of the three subgroups in the observed total flows, namely the primary, return and onward migration.

Therefore, these inherent data limitations may influence the analysis in this thesis in two ways. First, adopting the lifetime migration definition of the Census ignores multiple migration events that may have conducted by interprovincial migrants. For example, an interprovincial migrant originating from Province A now has stayed in Province B for more than six months, when the Census data collection takes place. But this migrant had conducted long-term stay (more than six months) in Province C before moving to Province B. The short-form data in the Census will only record the interprovincial move conducted by this migrant from Province A to Province B by ignoring the fact that this migrant had conducted interprovincial migration first from Province A to Province C and then from Province C to Province B. Second, such a lifetime migration definition is likely to under-count return movers that happen to carry out their return migration at the time of Census data collection. For instance, a migrant worker moves away from another province where he or she has stayed over six months and goes back to his or her home province when the Census data collection takes place at his or her origin; the Census data collector will first check that this migrant worker already takes residence in the registered Hukou address, and then the Census data collector will enter 'non-migrant' under this return migrant's name in the database. In summary, such data collection practices in the Census systematically overlook return and multiple migration events that have previously crossed the provincial administrative boundaries.

A popular method to deal with the under-count migration events in the Census is to power up the lifetime migration counts based on expert knowledge (Bell, Blake et al. 2002). For instance, Fielding (2010) powered up the lifetime migration velocity in China's 2000 Census by 2.2617 so as to make the comparison with the velocity of the five-year migration between 1995 and 2000. However, all the interprovincial migration counts modelled in this thesis are between six months and lifetime. It is therefore unlikely to generate different counts for the same flow due to time periods of different length in staying at the destination province, although in theory there are multiple different ways to collect migration flow data such as the eight different accounts of interprovincial migration in the long-form data of the 2000 Census. Most importantly, this thesis bases the analysis primarily on the interprovincial movers captured in the Census, rather than the accurate counts of moves. This thesis therefore does not apply any exponential index to power up the number of migrants contained in interprovincial flows. Nevertheless, care is still needed when interpreting the results, which are based on cross-sectional (between six months and lifetime) movers captured in the 2000 and 2010 Census rather than accurate longitudinal migration events over a certain period of time[36].

[36] More discussion is presented in Section 7.3 about how the way the data are measured and recorded may have influenced the analysis that follows.

Figure 3.1 Provincial administrative units in China[37]

This thesis focuses on the analysis of all the provincial units in China except for Taiwan which maintains its own system of population management (Figure 3.1), due to complicated historical and political reasons (Moeller 1994; Lin 2016). In other words, this thesis examines 31 provincial units in total, which are also collectively referred to as 'mainland China'. Every province of mainland China consists of several prefecture-level cities (地级市 *diji shi*), which are an administrative level lower than the province. On average, there are about 10 prefecture-level cities in each province. However, unlike the typical meaning of 'city', a prefecture-level city in mainland China not only contains municipal areas ('urban core') but also has governing rights towards surrounding counties (Figure 3.2). A prefecture-level city usually contains 10 counties, which in turn have governing responsibility towards towns,

[37] This map is created based on the geographical information provided by Institute of Geographic Sciences and Natural Resources research (IGSNRR), Chinese Academy of Sciences.

townships and villages within their boundaries respectively (Figure 3.2). In other words, each province contains both rural and urban areas that can be described with a conceptual diagram 3.2.

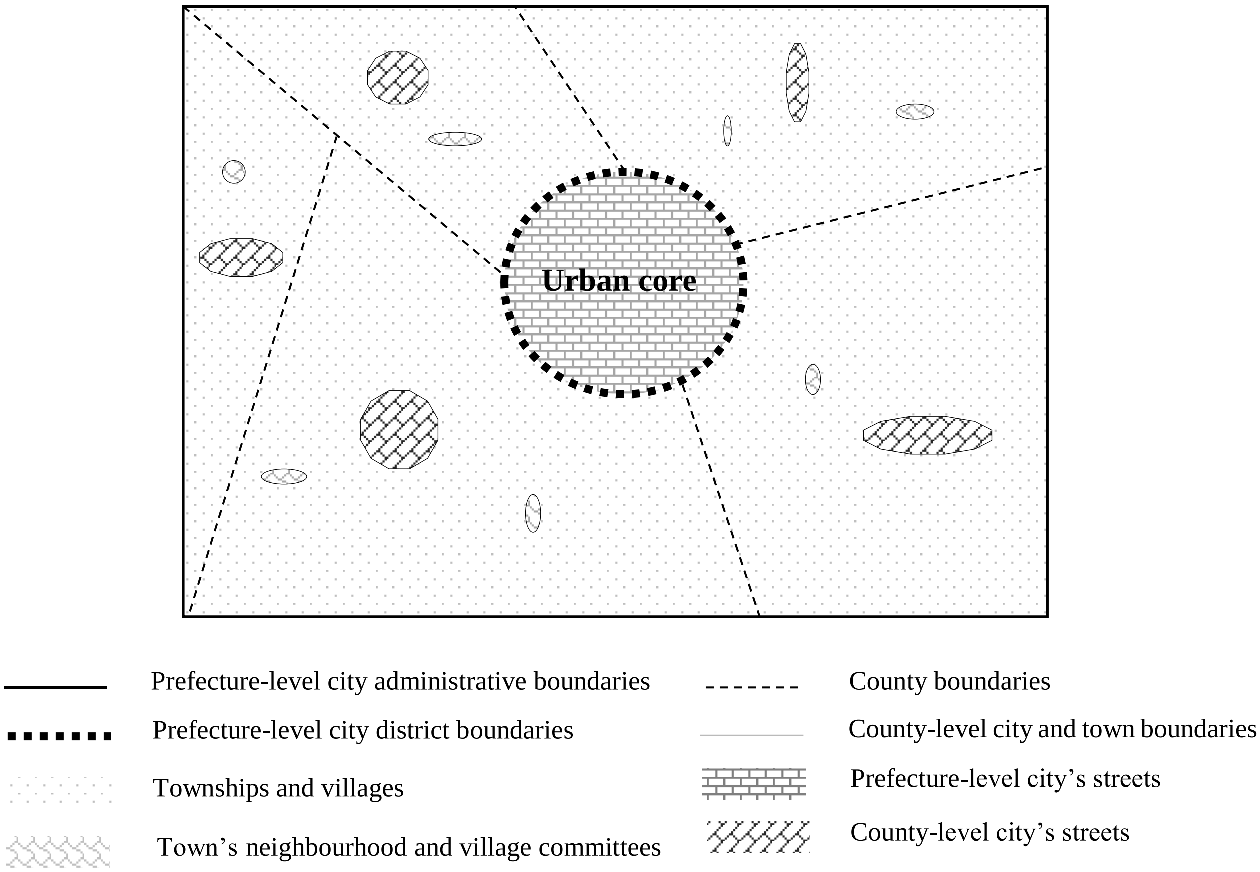

Figure 3.2 Conceptual definition of the administrative boundaries of rural and urban areas in a Chinese city

There are two mainstream classification criteria regarding the definitions of rural and urban areas (Chan and Zhang 1999). The first classification criterion defines rural and urban areas in a province based on the centre of Hukou registration (户口所在地 *hukou suozaidi*) (Chan and Zhang 1999), which is the definition adopted by the Census since 2000[38]. The Hukou status is binary and could only be either

[38] This rural and urban classification comes from the governmental regulation ‘Statistical classification of rural and urban areas (Beta version)’ (关于统计上划分城乡的规定（试行）, *guanyu tongjishang huafen chengxiang de guiding (shixing)*), which came into force on December 6th 1999. Its finalised version was approved by State Council of the People’s Republic of China in July 2008 and implemented nationwide since August 1st of 2008.

'agricultural' or 'non-agricultural' (Chan and Zhang 1999). To be more specific, urban areas include (a city's) 'Street' (街道 *jiedao*) and 'Neighbourhood committees of the town' (镇 *zhen*), whereas rural areas consist of 'Township' (乡 *xiang*) and 'Village committees of the town' (村 *cun*) in the Census (Figure 3.2). This classification criterion primarily applies to the collection of socio-economic data, without changing the administrative structure or relationship of internal political affairs[38]. The second classification criterion is based on administrative geographical boundaries (Chan 2007). There are various classifications of rural and urban areas under this criterion, depending on the level of administrative geographical boundaries adopted. Among all competing definitions based on administrative geographical boundaries, one very important and popular classification treats the 'urban core' as the 'urban area' (Chan 2007), whilst the rest is classified as the 'rural area' (Figure 3.2). The main reason is that many prefecture-level cities in China have an administrative area far larger than the urbanised area (Chan 2007), whilst the 'urban core' is regarded as highly urbanised and has high-density built-up area. Apart from that, the second classification criterion also offers a much clearer cut between rural and urban areas, which can stand as an advantage in some calculation and modelling.

This thesis adopts the first classification criterion. It is not only to follow the definitions in the Census, but also because the first classification criterion is a better description of the rural-urban continuum reality. This is exemplified by the definition of 'Town' in Figure 3.2, which consists of both urban ('Neighbourhood committees of the town') and rural ('Village committees of the town') areas. Whilst the binary conceptual simplification of space into rural and urban areas is necessary in extracting data and developing measurements for both criteria, recognising county-level cities and parts of towns as urban areas is more in line with the socio-economically lively and dynamic county-level cities and towns in China (Ma and Fan 1994; Eng 1997). Indeed, county-level cities and towns have played an important role in initiating and promoting bottom-up urbanisation in China (Ma 2002). The definition of the four interprovincial migration types is also based on the first classification criterion, which will be systematically examined in Chapter 4. For instance, interprovincial urban-urban migration is the movement made by migrants with urban Hukou between urban areas of different provinces, whilst interprovincial rural-urban migration is the movement made by migrants with rural Hukou from the origin province to urban areas of another province.

Table 3.1 Extract of the interprovincial migration flow matrix in 2000 (Unit: persons)

		Origin									
		Total	Beijing	…	Heilongjiang	…	Henan	…	Guangdong	…	Xinjiang
Destination	Total	42,418,562	91,702	…	1,174,048	…	3,069,955	…	430,446	…	156,263
	Beijing	2,463,217		…	90,014	…	334,605	…	22,987	…	10,892
	…	…	…	…	…	…	…	…	…	…	…
	Heilongjiang	386,641	1,437	…		…	14,946	…	1,627	…	465
	…	…	…	…	…	…	…	…	…	…	…
	Henan	476,239	3,278	…	11,226	…		…	8,905	…	7,313
	…	…	…	…	…	…	…	…	…	…	…
	Guangdong	15,064,838	11,356	…	50,306	…	1,005,219	…		…	13,116
	…	…	…	…	…	…	…	…	…	…	…
	Xinjiang	1,411,086	732	…	4,486	…	286,518	…	3,263	…	

Table 3.2 Extract of the interprovincial migration flow matrix in 2010 (Unit: persons)

		Origin									
		Total	Beijing	…	Heilongjiang	…	Henan	…	Guangdong	…	Xinjiang
Destination	Total	85,876,337	274,365	…	2,553,648	…	8,626,229	…	880,600	…	297,261
	Beijing	7,044,533		…	403,287	…	979,741	…	70,783	…	35,852
	…	…	…	…	…	…	…	…	…	…	…
	Heilongjiang	506,397	7,961	…		…	25,240	…	4,738	…	1,566
	…	…	…	…	…	…	…	…	…	…	…
	Henan	592,134	6,863	…	14,519	…		…	16,888	…	8,432
	…	…	…	…	…	…	…	…	…	…	…
	Guangdong	21,497,787	18,918	…	108,799	…	1,762,133	…		…	26,122
	…	…	…	…	…	…	…	…	…	…	…
	Xinjiang	1,791,642	2,263	…	10,887	…	386,615	…	11,933	…	

Table 3.1 and Table 3.2 each present an extract of the interprovincial migration matrix in 2000 and 2010 respectively, which pools together the separate rural and urban counts in each year. Apart from the provincial and national total, five provinces from all four regions are chosen to represent the scale and size of the interprovincial migration flow in mainland China (Figure 3.1): Beijing is the national capital and a province in the East region[39]; Guangdong is also a province of the East; Heilongjiang belongs to the Northeast region; Henan comes from the Central region; and Xinjiang is in the West region. Due to the huge population size in China, interprovincial migration flows are quite large and do not contain zero flows (Table 3.1 and Table 3.2). The total number of interprovincial migration is 42 and 86 million persons in 2000 and 2010 respectively (Table 3.1 and Table 3.2). On average, each flow contains 46 and 92 thousand of interprovincial migrants in 2000 and 2010 respectively (the full flow matrix of 31 provinces are presented in Table 3.6 and 3.7 of the appendix). The median count of interprovincial migration flows is 6,254 and 15,786 in 2000 and 2010 respectively. The largest flow in 2000 is from Hunan to Guangdong, which stands at 3 million; the smallest flow in the same year contains 43 people, originating from Tibet and ending at Liaoning and both provinces are very far apart. In 2010, the flow from Hunan to Guangdong still ranks the first with a huge number of migrants (about 5 million); the smallest flow is from Inner Mongolia to Tibet with 223 migrants.

The province with the largest out-migration is Sichuan in both 2000 and 2010, sending out 7 and 9 million interprovincial migrants in each year respectively (Table 3.3 and Table 3.4). Tibet remains as the province sending out the smallest number of interprovincial migrants in both 2000 (20 thousand) and 2010 (55 thousand) (Table 3.3 and Table 3.4). As with total in-migration, Guangdong ranks the first in both years, with in-flows containing 15 and 21 million in 2000 and 2010 respectively (Table 3.3 and Table 3.4). By contrast, Tibet ranks the last in both years, with in-flows totalling at 109 and 165 thousand in 2000 and 2010 respectively (Table 3.3 and Table 3.4).

[39] Map of the region boundary in China is presented in Section 6.3, as the role of region in the migration system is extensively discussed in Chapter 6.

Whilst in-migration and out-migration refer to the sum of all in-migration and out-migration interprovincial flows for each specific province respectively, net migration is calculated as the outcome of in-migration minus outmigration for that province (Table 3.3 and Table 3.4). In and out migration of Table 3.3 and Table 3.4 are the row and column totals in the previous tables 3.1 and 3.2 respectively. The in-migration, out-migration and net-migration rate refers to the proportion of in-migration, out-migration and net-migration among the total provincial population for each specific province respectively (Table 3.3 and Table 3.4). Such measurement considers the effect of provincial population size. Total migration is equal to the sum of in-migration and out-migration (Plane 1984). Migration effectiveness of a place is therefore calculated as the proportion of net-migration among the total migration of that place (Plane 1984), which can be denoted as the equation ((in-migration - out-migration) / (in-migration + out-migration)*100%).

Table 3.3 Interprovincial in-migration, out-migration, net migration and migration effectiveness in China in 2000

	2000						
	In-migration (row totals of Table 3.1)	Out-migration (column totals of Table 3.1)	Net-migration (in-migration - out-migration)	In-migration rate (%) (in-migration/total provincial population * 100%)	Out-migration rate (%) (out-migration/total provincial population * 100%)	Net-migration rate (%) (net-migration/total provincial population * 100%)	Migration effectiveness (%)((in-migration - out-migration) / (in-migration + out-migration)*100%)
Beijing	2,463,217	91,702	2,371,515	18.35	0.68	17.66	92.82
Tianjin	735,033	82,499	652,534	7.51	0.84	6.67	79.82
Hebei	930,455	1,218,975	-288,520	1.40	1.83	-0.43	-13.42
Shanxi	667,357	305,148	362,209	2.07	0.95	1.13	37.24
Inner Mongolia	547,923	504,557	43,366	2.38	2.19	0.19	4.12
Liaoning	1,045,165	361,944	683,221	2.52	0.87	1.64	48.55
Jilin	308,605	608,693	-300,088	1.16	2.29	-1.13	-32.71
Heilongjiang	386,641	1,174,048	-787,407	1.08	3.26	-2.19	-50.45
Shanghai	3,134,922	142,657	2,992,265	19.27	0.88	18.40	91.29
Jiangsu	2,536,889	1,715,634	821,255	3.50	2.37	1.13	19.31
Zhejiang	3,688,851	1,482,465	2,206,386	8.11	3.26	4.85	42.67
Anhui	230,116	4,325,830	-4,095,714	0.39	7.36	-6.97	-89.90
Fujian	2,145,256	810,576	1,334,680	6.35	2.40	3.95	45.15
Jiangxi	253,095	3,680,346	-3,427,251	0.63	9.20	-8.57	-87.13
Shandong	1,033,213	1,104,645	-71,432	1.15	1.23	-0.08	-3.34
Henan	476,239	3,069,955	-2,593,716	0.52	3.38	-2.85	-73.14
Hubei	609,733	2,805,187	-2,195,454	1.03	4.73	-3.70	-64.29
Hunan	348,838	4,306,851	-3,958,013	0.55	6.83	-6.27	-85.01

Guangdong	15,064,838	430,446	14,634,392	17.95	0.51	17.43	94.44
Guangxi	428,188	2,441,847	-2,013,659	0.98	5.59	-4.61	-70.16
Hainan	381,792	119,403	262,389	5.10	1.60	3.51	52.35
Chongqing	403,159	1,005,773	-602,614	1.33	3.31	-1.98	-42.77
Sichuan	536,246	6,937,793	-6,401,547	0.65	8.47	-7.82	-85.65
Guizhou	408,519	1,596,461	-1,187,942	1.17	4.58	-3.41	-59.25
Yunnan	1,164,402	343,542	820,860	2.79	0.82	1.97	54.44
Tibet	108,669	19,849	88,820	4.18	0.76	3.41	69.11
Shaanxi	426,029	804,454	-378,425	1.21	2.29	-1.08	-30.75
Gansu	227,888	585,868	-357,980	0.91	2.35	-1.43	-43.99
Qinghai	124,307	94,988	29,319	2.61	1.99	0.62	13.37
Ningxia	191,891	90,163	101,728	3.55	1.67	1.88	36.07
Xinjiang	1,411,086	156,263	1,254,823	7.71	0.85	6.85	80.06

Table 3.4 Interprovincial in-migration, out-migration, net migration and migration effectiveness in China in 2010

	2010						
	In-migration (row totals of Table 3.2)	Out-migration (column totals of Table 3.2)	Net-migration (in-migration - out-migration)	In-migration rate (%) (in-migration/total provincial population * 100%)	Out-migration rate (%) (out-migration/total provincial population * 100%)	Net-migration rate (%) (net-migration/total provincial population * 100%)	Migration effectiveness (%)((in-migration - out-migration) / (in-migration + out-migration)*100%)
Beijing	7,044,533	274,365	6,770,168	56.11	2.19	53.93	92.50
Tianjin	2,991,501	273,134	2,718,367	30.16	2.75	27.40	83.27
Hebei	1,404,673	3,498,253	-2,093,580	1.95	4.86	-2.91	-42.70
Shanxi	931,653	1,083,291	-151,638	2.68	3.12	-0.44	-7.53
Inner Mongolia	1,444,181	1,067,556	376,625	5.92	4.37	1.54	14.99
Liaoning	1,786,530	1,014,028	772,502	4.20	2.38	1.82	27.58
Jilin	456,499	1,372,853	-916,354	1.68	5.06	-3.38	-50.09
Heilongjiang	506,397	2,553,648	-2,047,251	1.32	6.68	-5.35	-66.90
Shanghai	8,977,000	250,340	8,726,660	63.28	1.76	61.52	94.57
Jiangsu	7,379,253	3,058,880	4,320,373	9.84	4.08	5.76	41.39
Zhejiang	11,823,977	1,853,940	9,970,037	24.98	3.92	21.06	72.89
Anhui	717,463	9,622,595	-8,905,132	1.05	14.02	-12.98	-86.12
Fujian	4,313,602	1,667,254	2,646,348	12.20	4.71	7.48	44.25

Jiangxi	599,942	5,787,395	-5,187,453	1.27	12.28	-11.01	-81.21
Shandong	2,115,593	3,095,717	-980,124	2.22	3.24	-1.03	-18.81
Henan	592,134	8,626,229	-8,034,095	0.57	8.27	-7.70	-87.15
Hubei	1,013,612	5,889,792	-4,876,180	1.64	9.54	-7.90	-70.63
Hunan	724,982	7,228,896	-6,503,914	1.02	10.21	-9.19	-81.77
Guangdong	21,497,787	880,600	20,617,187	25.28	1.04	24.25	92.13
Guangxi	841,806	4,184,566	-3,342,760	1.63	8.11	-6.48	-66.50
Hainan	588,463	275,751	312,712	6.93	3.25	3.69	36.18
Chongqing	945,194	3,506,899	-2,561,705	2.85	10.58	-7.73	-57.54
Sichuan	1,128,573	8,905,128	-7,776,555	1.25	9.90	-8.64	-77.50
Guizhou	763,294	4,048,596	-3,285,302	1.83	9.73	-7.90	-68.27
Yunnan	1,236,549	1,482,442	-245,893	2.71	3.25	-0.54	-9.04
Tibet	165,423	55,185	110,238	5.72	1.91	3.81	49.97
Shaanxi	974,362	1,960,598	-986,236	2.54	5.11	-2.57	-33.60
Gansu	432,833	1,593,265	-1,160,432	1.59	5.86	-4.27	-57.27
Qinghai	318,435	242,086	76,349	5.77	4.38	1.38	13.62
Ningxia	368,451	225,794	142,657	5.83	3.57	2.26	24.01
Xinjiang	1,791,642	297,261	1,494,381	8.85	1.47	7.38	71.54

Shanghai has the highest in-migration rate in both 2000 and 2010 (Table 3.3 and Table 3.4), standing at 19.27% and 63.28% respectively. Other provinces such as Beijing (18.35% in 2000 and 56.11% in 2010) and Guangdong (17.95% in 2000 and 25.28% in 2010) also have relatively high in-migration rates in both years (Table 3.3 and Table 3.4). Those provinces are in the East region and relatively more developed and urbanised, due to the fact that they have enjoyed long-term preferential development policies (Fan 1995; Démurger, Sachs et al. 2002; Fan, Kanbur et al. 2009). By comparison, Anhui has the lowest in-migration rate (0.39%) in 2000, whilst Henan has the lowest in-migration rate (0.57%) in 2010 (Table 3.3 and Table 3.4). Importantly, both Anhui and Henan are in the Central region, which is already famous for its high population density and low urbanisation level (Fan 2005a).

As with out-migration rate, Jiangxi ranks the first in 2000 with an out-migration rate of 9.20% (Table 3.3), whilst Anhui has the highest out-migration rate (14.02%) in 2010 (Table 3.4). Similar to Anhui, Jiangxi is in the Central region too and suffers from high population density and low urbanisation level (Fan 2005a). Other Central-region provinces such as Hunan (6.83% in 2000 and 10.21% in 2010) also have high out-migration rates (Table 3.3 and Table 3.4). By contrast, Guangdong has the lowest out-migration rates in both years (0.51% in 2000 and 1.04% in 2010). Other East-region provinces such as Shanghai (0.88% in 2000 and 1.76% in 2010) and Beijing (0.68% in 2000 and 2.19% in 2010) also have very low out-migration rates. All of this points to the same fact that provinces of the East region are relatively more developed and urbanised (Fan 1995; Démurger, Sachs et al. 2002; Fan, Kanbur et al. 2009).

The net-migration rate measures the contribution of net-migration that results in population change, considering the effect of provincial population size. East-region provinces such as Beijing, Shanghai and Guangdong have high net-migration rates in both 2000 and 2010. For instance, Shanghai (18.40% in 2000 and 61.52% in 2010) and Beijing (17.66% in 2000 and 53.93% in 2010) have the highest and second highest net-migration rates respectively. Jiangxi of the Central region and Sichuan of the West region have the lowest and second lowest net-migration rates respectively in 2000; two Central-region provinces Anhui and Jiangxi have the lowest and second lowest net-migration rates respectively in 2010. These observations of net-migration rates echo with findings of former literature that provinces of the East region are relatively more developed and urbanised than provinces of the Central and the West region (Fan 1995; Démurger, Sachs et al. 2002; Fan, Kanbur et al. 2009).

The calculation of migration effectiveness does not consider the population size of either the origin or the destination. Instead, it measures the percentage of migration turnover contributing to population change (Plane 1984). In the context of China, the national migration effectiveness of interprovincial migration flows is calculated as (Sum of (absolute value of (in-migration - out-migration) for each province) / (Sum of (value of (in-migration + out-migration) for each province))*100%), whilst interprovincial migration effectiveness of a specific province is calculated as ((in-migration - out-migration) / (in-migration + out-migration)*100%) on the provincial level (Plane 1984). The national migration effectiveness of interprovincial migration flows reflects the structural shift in the national interprovincial migration system, particularly representing the regional and urban hierarchic shifts of capital and labour (Plane 1984). Based on Table 3.3 and 3.4, the national migration effectiveness of interprovincial migration flows is 67.56% and 68.78% in 2000 and 2010 respectively, meaning that 67.56% and 68.78% of interprovincial migration in mainland China result in net provincial population gain or loss in 2000 and 2010. This also shows that interprovincial migration movements become slightly more unidirectional rather than bi-directional from 2000 to 2010 (Fan 2005b). The slight increase in the national effectiveness of interprovincial migration flows from 2000 to 2010 fundamentally reflects the deepening industrial development process in China during this period (Plane 1984). In other words, the high values in both years may imply that China was in the stage of urban system growth and expansion with rising rural-to-rural migration between 2000 and 2010 (Zelinsky 1971), which means that it still has a short way to go before commencing substantial post-industrialism regional and urban hierarchic shifts of capital and labour by replacing rural-to-urban migration with urban-to-urban migration[40] (Zelinsky 1971; Plane 1984).

Due to the different calculation procedure of national and provincial migration effectiveness, the interpretation of the interprovincial migration effectiveness for a specific province is different from that of the national migration effectiveness of interprovincial migration flows[41]. In 2000, many East-region provinces such as Beijing (92.82%), Shanghai (91.29%), Tianjin (79.82%) and Guangdong (94.44%) and one West-region province Xinjiang (80.06%) have very high migration effectiveness, meaning that (positive) migration turnover has contributed substantially to population change. The high migration effectiveness of East-region provinces is related to the preferential regional development policies (Fan 1995; Démurger, Sachs et al. 2002; Fan, Kanbur et al. 2009), whilst Xinjiang's high migration effectiveness may lie in the long-standing state policy in encouraging internal migration into it (Becquelin 2000; Howell and Fan 2011). Central-region provinces such as Anhui (-89.90%), Jiangxi (-87.13%), Hunan (-85.01%) and Henan (-73.14%), and one West-region province Sichuan (-85.65%) have very low migration effectiveness, meaning that (negative) migration turnover has contributed substantially to population change in these provinces. The low migration effectiveness of both Central-region provinces and Sichuan of the West region is related to their low urbanisation and development level (Fan 1995; Démurger, Sachs et al. 2002; Fan, Kanbur et al. 2009).

In 2010, many East-region provinces such as Beijing (92.50%), Shanghai (94.57%), Tianjin (83.27%) Guangdong (92.13%) and Zhejiang (72.89%) and one West-region province Xinjiang (71.54%) still have very high migration effectiveness. Similar to 2000, (positive) migration turnover also has contributed substantially to population change in 2010. The only notable change is the migration

[40] See sub-section 2.3.1.4 for more details.
[41] For instance, if all migrants are in-migrants (out-migration equals zero) for a province, it will give a migration effectiveness value of 100% for that province. By contrast, a provincial migration effectiveness value of -100% means that all migrants are out-migrants and there are no in-migrants for that specific province. Additionally, a province with equal numbers of in-migrants and out-migrants and no net-migration would have a provincial migration effectiveness value of zero, meaning that migration essentially does not contribute to population change (or at least the population size change) within that province.

effectiveness of Zhejiang, which increases from 42.67% to 72.89% between 2000 and 2010. Central-region provinces such as Anhui (-86.12%), Jiangxi (-81.21%), Hunan (-81.77%) and Henan (-87.15%), and one West-region province Sichuan (-77.50%) still have very low migration effectiveness in 2010. Between 2000 and 2010, there is no substantial change in the (negative) migration turnover of these provinces by comparison.

This thesis makes an important distinction between two conceptions of migration flows - ‘flow/ counter-flow’ and ‘flow-pairs’, though they have been usually used synonymously in the literature (Ravenstein 1885; Molho 2013). To illustrate, in the literature ‘flow’ and ‘counter-flow’ refer to a pair of migration flows between two places (for instance, Province A and Province B); if the flow originates from Province A and ends at Province B, its counter-flow is regarded as originating from Province B and ending at Province A, whereas a ‘flow-pair’ refers to a pair of flows with opposite directions between two places. In this thesis, however, they are defined differently to better represent the relationship and direction of the four interprovincial migration types. For example, the urban-rural migration flow from Province A to Province B, and the migration flow of the same type but with the opposite direction (i.e., from Province B to Province A) are defined as the flow-pairs of urban-rural migration between Province A and Province B (upper sub-plot of Figure 3.3). Conversely, the urban-rural flow from Province A to Province B, and the rural-urban flow of the opposite direction (i.e., from Province B to Province A) are defined as the flow/ counter-flow between urban areas of Province A and rural areas of Province B (lower sub-plot of Figure 3.3).

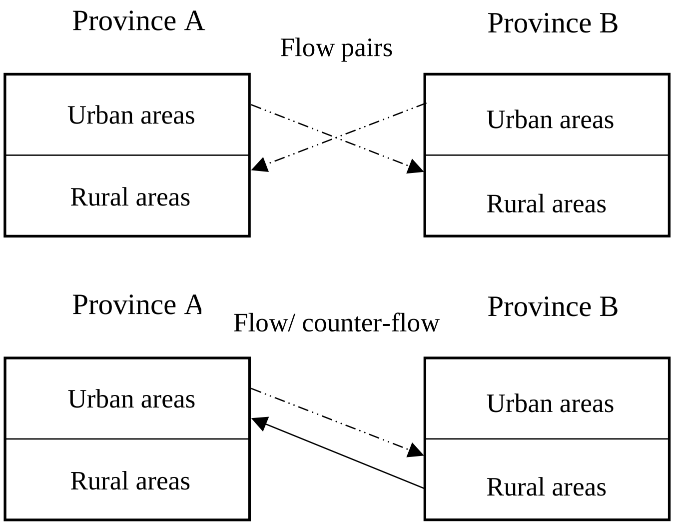

Figure 3.3 Conceptions of ‘flow-pairs’ and ‘flow/ counter-flow’ for urban-rural migration streams

Between ‘flow/ counter-flow’ and ‘flow-pairs’, flow-pairs are the major focus of this thesis, and this definition could facilitate the analysis of the spatial structures of flows within each stream type. The point of doing this is also to avoid getting into the situation where the flows between urban and rural areas must necessarily trade-off against each other, as the research hypothesis is that different streams are fundamentally different from each other. This is also because the pairwise relationship is an inherent feature of migration flows, which can be summarised as positive ‘reciprocity’. The positive ‘reciprocity’ of migration flows means that if one province sends a higher than predicted number of migrants to another, it is likely to see a higher than predicted number of migrants in the return direction. Reciprocity in flows between provinces can be an important feature to be tested, representing the correlation of flow-pairs. The emphasis on ‘flow-pairs’ in this thesis thus reflects the importance of dyads[42] in data analysis. Dyads are a type of paired data, in which there are two individual research units of interest linked by certain networks or processes (Gonzalez and Griffin 1997; Gonzalez and Griffin 2012).

[42] Dyadic data analysis has its origins in psychology, in which pairwise ties of subjects have been the prime research interest in a large body of studies (Campbell and Kashy 2002, Kenny and Kashy 2011a, Gonzalez and Griffin 2012). Dyads have gained popularity outside psychology such as in politics (Maoz 2004), management (Knight and Humphrey), gender relationships (Thacker and Ferris 1991), and the study of other socio-economic activities (Koster and Leckie 2014).

Therefore, dyadic data analysis also fits well with interprovincial migration flow data and thus the definition of ‘flow-pairs’ has been adopted in this thesis.

3.1.2 Populations and incomes

In estimating interprovincial migration flows, population and income predictors are specified as their rural and urban components drawn from the Census and the China Statistical Yearbook respectively. Following the standard of the China Statistical Yearbook, per capita annual disposable income of urban households and per capita annual net income of rural households are used to measure the level of wealth in rural and urban areas of each province. Rural and urban components of the provincial population are taken from the short-form dataset for the whole population contained in the census, which exclude counts of the migrants.

Figure 3.4 shows the urban population of provinces in 2000 (left) and 2010 (right). Provincial urban population is clustered into four different groups based on the Jenks natural breaks classification method, which seeks to minimise the variance within groups and maximise the variance between groups (Jenks and Caspall 1971). The distribution pattern of provincial urban population is similar overall in 2000 and 2010. For instance, the number of provinces with the lowest urban population (Hainan, Tibet, Qinghai and Ningxia) remains the same in 2000 and 2010. The number of provinces with the second lowest urban population almost remain the same during this period, with the only notable change of Jiangxi degrading from group of the second largest urban population in 2000 to group of the second lowest urban population in 2010. Heilongjiang, Henan and Hubei decrease from group of the largest urban population in 2000 to group of the second largest urban population in 2010.

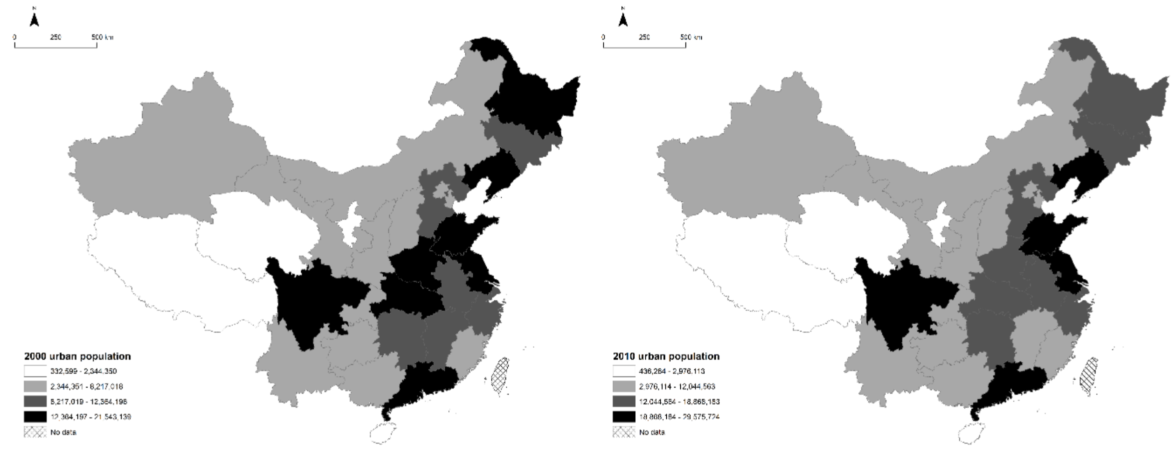

Figure 3.4 Urban population of provinces in 2000 and 2010 (unit: persons)

Similarly, Figure 3.5 shows the four groups of provincial rural population in 2000 (left) and 2010 (right) based on the Jenks natural breaks classification method. The overall distribution pattern of provincial rural population is almost the same in 2000 and 2010. The number of provinces with the lowest rural population (Hainan, Tibet, Qinghai and Ningxia) remains the same in 2000 and 2010, and so is the number of provinces with the second lowest rural population (Heilongjiang, Inner Mongolia, Jilin, Liaoning, Shanxi, Gansu, Xinjiang and Chongqing) during this period. The only noticeable change is

Anhui, which degrades from group of the largest rural population in 2000 to group of the second largest rural population in 2010.

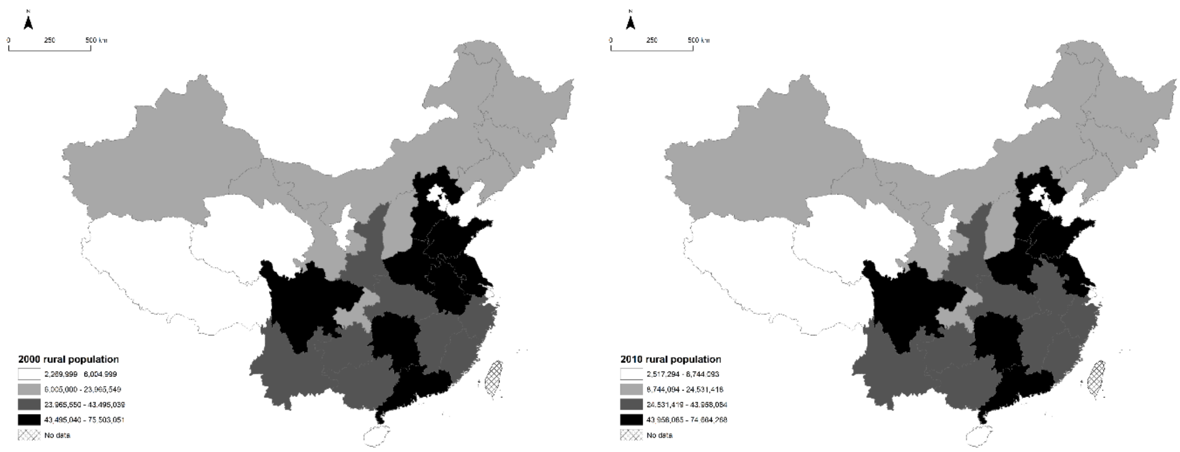

Figure 3.5 Rural population of provinces in 2000 and 2010 (unit: persons)

The overall distribution of provincial urban income experiences substantial changes between 2000 (left) and 2010 (right) (Figure 3.6). Similar to figures of provincial rural and urban population, the four groups of provincial urban income in 2000 and 2010 are also based on the Jenks natural breaks classification method. All four groups change remarkably. There are fifteen provinces with the lowest urban income (Heilongjiang, Inner Mongolia, Jilin, Liaoning, Shanxi, Henan, Anhui, Jiangxi, Hainan, Guizhou, Hubei, Shaanxi, Ningxia, Gansu and Qinghai) in 2000. But only four provinces (Heilongjiang, Guizhou, Gansu and Qinghai) of them remain in the same group in 2010 whilst the rest of the eleven provinces upgrade to group of second lowest urban income. Xinjiang degrades from group of the second lowest urban income in 2000 to group of the lowest urban income in 2010, and Tibet degrades from group of the second largest urban income in 2000 to group of the second lowest urban income in 2010. Similarly, Guangdong degrades from group of the largest urban income in 2000 to group of the second largest urban income in 2010, whilst Shandong and Jiangsu upgrade from group of the second lowest urban income in 2000 to group of the second largest urban income in 2010.

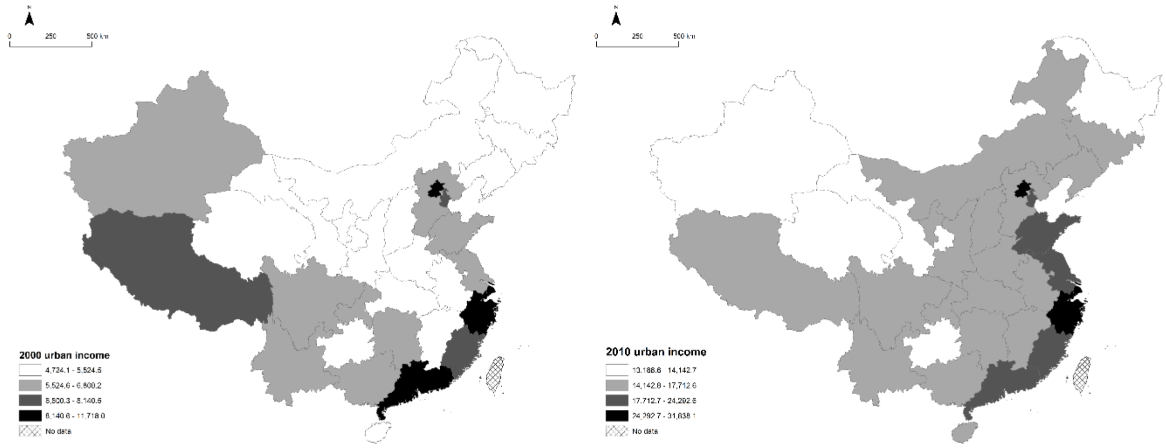

Figure 3.6 Urban income of provinces in 2000 and 2010 (unit: yuan)

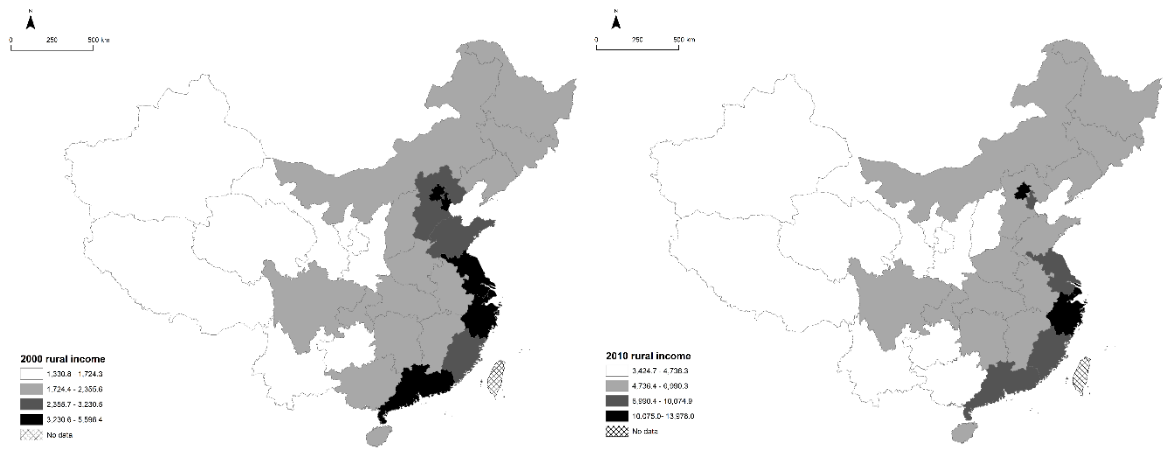

Figure 3.7 Rural income of provinces in 2000 and 2010 (unit: yuan)

Compared with the distribution of provincial urban income, the overall distribution of provincial rural income almost remains the same between 2000 (left) and 2010 (right) (Figure 3.7). Like former figures, the four groups of provincial rural income in 2000 and 2010 are also based on the Jenks natural breaks classification method. There are a few noticeable changes during this period. Specifically, Tianjin, Jiangsu and Guangdong degrade from group of the largest provincial rural income in 2000 to group of the second largest provincial rural income in 2010. The rest of the provinces remain in their former group.

The income inequality between provinces is represented by incomes of rural and urban areas in this thesis, which is the geographic distribution difference of income in rural and urban areas. Conventionally, such income inequality can also be measured as Gini coefficients (Chang 2002) or income gaps[43] (Kanbur and Zhang 1999; Sicular, Ximing et al. 2007). There are two major reasons for not using Gini coefficient or income gaps. First, this thesis focuses on investigating the multi-directional effects of rural and urban incomes[44], which can be straightforwardly specified as the rural and urban components of provincial income. Second, replacing the rural and urban components of provincial income with their four income gaps (rural-rural, rural-urban, urban-rural and urban-urban) will inevitably bring about collinearity issues in the model[45]. On account of these two reasons, this thesis chooses to specify between-province income inequality as the rural and urban components of provincial income.

Whilst the reasons to use the rural and urban components of provincial income have been extensively addressed, there are also various theoretical justifications to include the rural and urban components of the population. The origin population represents the potential for the origin to produce migrants (King 2012; Laczko, Tjaden et al. 2017), and thus the rural and urban components of the origin population measure how much potential the origin province has in exporting rural and urban migrants respectively. The destination population is an effective proxy for expected employment opportunities, consumer markets and other services (Ravenstein 1885; Ravenstein 1889; Dorigo and Tobler 1983; Massey, Arango et al. 1993; Taylor 1999). Therefore, the rural and urban components of the destination

[43] There are two major ways to measure the rural and urban income gap in the literature – the ratio of one income (urban income, for instance) to the other (for example, rural income), which is known as the relative gap and most popular in the literature due to its flexibility in handling both cross-sectional and longitudinal data sets (Kanbur and Zhang 1999, Sicular, Ximing et al. 2007); and the lesser popular measurement of the absolute gap, defined as the difference between urban and rural mean incomes.
[44] See Sub-section 2.1.2 for more detail.
[45] For example, if rural-rural, rural-urban and urban-rural relative income gaps are entered in the model, the urban-urban relative income gap will be perfectly collinear to them and no further investigation regarding the urban-urban income gap will be given.

population measure how attractive the rural and urban areas of the destination province are to in-migrants.

3.1.3 Distance

Migration is a key element of population dynamics (Gobermeyers 1978), and a migration process could be defined differently according to the migratory distance and the type of boundaries crossed by the migrant in the move (Levy 2010; Otoiu 2014). For instance, the major focus of this thesis is interprovincial migration, which is distinguished from interprovincial migration based on whether the involved internal migrants have crossed the provincial boundaries or not during the move. Distance is rated as one prevailing interest among empirical migration studies[46] (Zelinsky 1971; Gobermeyers 1978; Meentemeyer 1989). In the context of China, for example, Shen (1996) noticed that most large interprovincial migration flows occurred between neighbouring provinces in mainland China and thus evidenced the sensitivity of migration to distance in 1987 and 1990.

The migratory distance is also closely relevant to the choices of destinations and the spatial patterns of migration flows. For instance, by studying the migratory distance of interprovincial flows, Zhu, Ma et al. (2005) found that destination choices of migrants clustered on three coastal municipality regions (i.e. Beijing-Tianjin-Dalian, Shanghai-Nanjing-Hangzhou and Guangzhou-Shenzhen-Xiamen) in 1996. Echoing these findings, Bao, Shi et al. (2005) also found that coastal cities (except for Xinjiang) were the centres of destination preferences for interprovincial migrants before 2000, regardless of how faraway the origin places were. Such coastal-region-oriented spatial patterns were repeatedly found among both skilled and unskilled labour migrants before 2010 (Liu and Shen 2014b; Liu, Stillwell et al. 2014).

[46] Zelinsky (1971) summarised the interpretation of distance in migration research as follows: ‘Throughout the migrational literature, space is almost always treated as an absolute, with distances between points reckoned as constant. Although this is valid in a physical sense, it is misleading in any functional approach to space’ (Zelinsky 1971, p. 226).

The migratory distance, also known as the origin-destination distance, is calculated as the direct-line distance between provincial capital cities in this thesis. The direct-line distance follows the conventional measurement of Cartesian distance in migration studies (Brown, Odland et al. 1970; Courgeau and Baccaini 1989). Its alternative measurements include transportation (such as railway) distance and travelling cost and time (Sjaastad 1962; Molho 2013; Thomas, Stillwell et al. 2015). For instance, highway distances are occasionally used as an explanatory variable to estimate the volume of interprovincial migration flows (Shen 1999). In studying internal migration in China, however, the direct-line distance between the origin and the destination have been a very popular measurement of the migratory distance and are widely used in gravity and other models (Zhang and Song 2003; Han, Hayashi et al. 2009). Among all alternative measurements, the railway transportation distance has also been frequently adopted as a proxy measurement of the migratory distance for between-province movements (Chan, Liu et al. 1999; Fan 2005b; Shen 2016a). In addition, the train travel time is used to estimate migrants' sensitivity towards the migratory distance (Li 2004). The reason behind using train travel time or distance lies in that railway travelling is believed to be the major means of long-distance between-province travel and thus most relevant to migrants, due to its relatively cheap cost compared with flights (Fan 2005b). However, it is found that there are minimal differences of the results using Cartesian distance or railway distance in the model estimations of interprovincial migration flows (Fan 2005b). This thesis therefore will focus on Cartesian distance and not further pursue alternative distance measurements.

This thesis chooses the direct-line measurement of distance also because it has a straight-forward interpretation of distance decay, apart from the fact it is a proven effective measurement in the established literature (Levy and Wadycki 1974; Greenwood 1985; Fan 2005b). The linear and non-linear effects upon migration are examined in the following results chapters. The linearisation of distance decay follows the typical and direct interpretation of the gravitation law by holding distance as inversely proportional to migration (Courgeau and Baccaini 1989; Fan 2005b; Thomas, Stillwell et al. 2015). This is also in line with Tobler's first law of geography, as migrants are more likely to move to near rather than faraway places in general (Tobler 1970), controlling all other migratory factors (Shaw 1975). Besides specifying distance with typical non-linear functions, the exploration of non-linear effects also relies on examining the contiguity of provinces and regions. This is because unbalanced regional development policies have led to segmented provincial and regional labour markets, wherein certain migration flows are favoured whilst others are discouraged.

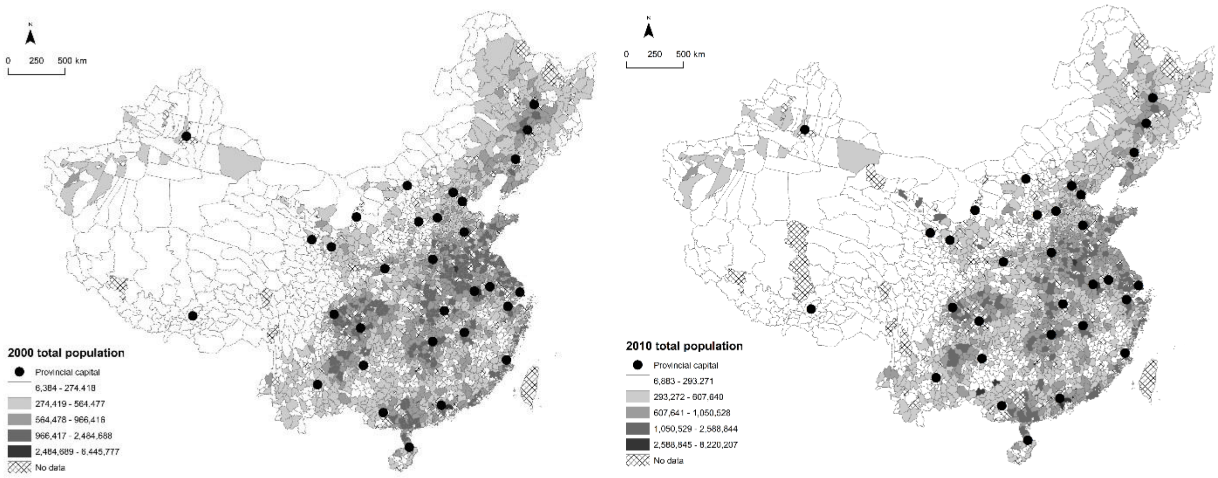

Figure 3.8 County-level total population distribution in 2000 and 2010[47]

Choosing the distance between provincial capitals as a representative measurement of the actual migratory distance could also be justified by county-level population distribution maps in 2000 and 2010 (Figure 3.8). Because of the continuous growth of cities and the frequent adjustment of county-level boundaries, Figure 3.8 only depicts counties without any boundary changes[47], accounting for about 90% of all the counties in both 2000 and 2010 respectively. As shown by Figure 3.8, provincial capitals usually have the largest population within the province and tend to be centres of provincial population distribution. Therefore, the distance between provincial capitals can be used as an effective proxy measurement for the between-province migratory distance, especially when the individual migrant's

[47] Population data are from Census 2000 and 2010. The base-map of Chinese counties adopts the county-level administrative boundaries in 2008, due to limitation of data. The county-level geographical data is provided by Data Center for Resources and Environmental Sciences (RESDC), Chinese Academy of Sciences.

migratory distance is unknown in the Census. This is also in line with the interprovincial migration reality in China. In fact, the distance between provincial capitals has been widely adopted in the interprovincial migration studies in China (Fan 2005b; Liu, Qi et al. 2015).

As provinces and regions usually implement distinctive migration management policies, the origin-destination distance could also reflect such effects to an extent. Recalling the discussion of the Hukou system[48], migrants are usually faced with certain regulations from the Hukou system during the relocation process. Given the same origin-destination distance between two provinces, the four interprovincial migration types could therefore be affected by different extents of resistance from the system. In practical terms, the research interest here is the marginal effect of distance on migration (Grinblatt and Keloharju 2001; Ai and Norton 2003). That is, by how much migration drops for a 1 km increase in distance. Consequently, it is of practical significance to find whether the distance affects the four migration types differently, particularly as the effect of migration management policies is difficult to measure directly.

[48] See Sub-section 2.3.1.2 of Chapter 2.

Table 3.5 Covariates of each research chapter

Chapter	Dependent variable	Unit	Independent variable	Unit	Source
			origin-destination distance	000s m	IGSNRR[49]
			origin urban population	000,000s persons	Census
	total migration	000s persons	origin rural population	000,000s persons	Census
	urban-rural migration	00s persons	origin urban income	000s yuan	Yearbook
4	rural-urban migration	00s persons	origin rural income	000s yuan	Yearbook
	rural-rural migration	00s persons	destination urban population	000,000s persons	Census
	urban-urban migration	00s persons	destination rural population	000,000s persons	Census
			destination urban income	000s yuan	Yearbook
			destination rural income	000s yuan	Yearbook
			origin-destination distance	000s m	IGSNRR[49]
			origin urban population	000,000s persons	Census
5	urban-rural migration	00s persons	origin urban income	000s yuan	Yearbook
			destination urban population	000,000s persons	Census
			destination urban income	000s yuan	Yearbook
			origin-destination distance	000s m	IGSNRR[49]
			origin urban population	000,000s persons	Census
			origin urban income	000s yuan	Yearbook
			destination urban population	000,000s persons	Census
6	urban-rural migration	00s persons	destination urban income	000s yuan	Yearbook
			neighbouring province	N/A	IGSNRR[50]
			provincial region	N/A	NDRC[51]
			interactions of neighbouring province and provincial region	N/A	N/A

Table 3.5 summarises covariates used in each research chapter and their units and sources. Chapter 4 first examines total interprovincial migration in both 2000 and 2010 to set the overall scene for this thesis; it then investigates the four migration flow types (i.e. urban-rural, rural-urban, rural-rural, and urban-urban migration) in 2010 to answer the first research question[52], which is relevant to the unavailability of four-flow data in 2000. In examining interprovincial flows, Chapter 4 considers the

[49] Distance is calculated based on the geographical information provided by IGSNRR.
[50] Neighbouring province is calculated based on the neighbouring relationship of provinces in the administrative map.
[51] Provincial region is calculated based on the definition of National Development and Reform commission (NDRC), China.
[52] See Section 1.2 of Chapter 1 for more details.

competition of within- and between-provincial places in attracting migrants by specifying both population and income covariates with their rural and urban components, in order to better understand associations between interprovincial migration and regional inequality on the rural and urban level. Chapter 5 chooses urban-urban migration flow in 2010 as the response variable, because urban-urban migration flow is under-researched and it could best represent features of other three flow types. Chapter 5 does not use the full data set of the independent variables, as its purpose is to present the proposed extended model in as simple an accessible form as possible. This choice is also due to the limit of time and resources in doing this PhD. Chapter 6 further expands the investigation to the continuity of provincial and regional boundaries by adding new covariates on the basis of Chapter 5's covariate selection.

Finally, it should be emphasised again that this thesis is primarily a cross-sectional study, although interprovincial migration in 2000 and 2010 is compared in Chapter 4. Cautions should be therefore exercised when reading this thesis. This is because other factors associated with migration except for incomes, populations and distance are unexplored and could change with time. To be more specific, these unexplored factors might vary from one point to another, although the overall socio-economic conditions under which interprovincial migration occurs have been largely taken into account in the models. More detailed discussion about the potential impact from this could be found in Sub-section 7.3 of Chapter 7.

3.2 Methodology

This section starts by justifying choice of the gravity model in the thesis and discussing the alternative modelling techniques. It then introduces the traditional form of the gravity model of migration, and subsequently presents its traditional and general form in linear regression formulation.

This thesis chooses the gravity model of migration to investigate the interprovincial migration in China, owing to its flexibility in incorporating origin and destination factors when the migration flow is modelled (Peeters 2012; Flores, Zey et al. 2013; Fitzgerald, Leblang et al. 2014; Poprawe 2015; Shen 2015). Although the gravity model of migration is widely regarded as a good tool and the most popular model analysing migration flows, studies using it to study China's internal migration are relatively scarce (Fan 2005b). Particularly, there is a lack of systematic investigation upon associations between interprovincial migration flows and regional inequality on rural and urban, province and region levels in China using gravity models.

Other alternative methodological frameworks of modelling migration flows include stochastic population growth model (Engen 2007), cellular automata model (Dabbaghian, Jackson et al. 2010), singularly perturbed population model (Banasiak and Goswami 2015), and Bayesian model for migration decisions (Wilber 1965). However, some of these alternative methods do not fit well with the chosen data and measurements due to requirement of more or finer-scale data, such as singularly perturbed population model, stochastic population growth model, and Bayesian model for migration decisions. Other alternative methods could not fully answer the research questions, such as cellular automata model or other micro-simulation methods.

3.2.1 The gravity model of migration

Let m_{ij} denote the number of migrants who move from origin province i ($i = 1, \ldots, n$) to destination province j ($j = 1, \ldots, n$). The traditional gravity model of migration can then be written as

$$m_{ij} = k \frac{p_i^a p_j^b}{d_{ij}^c} \tag{3.1}$$

where p_i and p_j denote the populations of province i and j, d_{ij} denotes the distance between them, k is a constant, and a, b and c are the powers originally hypothesised to take the values 1, 1 and 2, respectively, resembling Newton's law of gravity (Christian and Braden 1966; Claeson 1969). The underlying assumption is that a migration flow between two places is determined by the attraction between the origin and the destination, which can be measured as proportional to the product of the 'masses' of two places (the origin and destination populations in this case) and inversely proportional to the square of the distance between them. However, this assumes that $m_{ij} = m_{ji}$ whereas, in practice, the bilateral migration flows between pairs of places are often not equal to each other in the number of migrants that they contain. Furthermore, there is no reason why 'mass' should be measured only by population size. Thus the original form of the gravity model was soon refined and the artificial constraints on power coefficients of a, b and c were then released (Wilson 1971; Ginsberg 1972).

3.2.2 The traditional linear regression formulation

The mathematical form of the gravity model (Equation 3.1) can be taken logs on both sides of the equation and thus re-expressed as a multiple regression equation (Equation 3.2). This is extremely useful as now the gravity model can be treated as a statistical model and it thus enables the estimation of the model parameters from the data (Flowerdew 2010). Standing as a multiple regression equation, it is also now possible and flexible to introduce additional predictors (other than populations of origin and destination and the migratory distance) into the original gravity model (Field 2013). Moreover, the log-transformation of the original gravity model essentially leads to a linear regression model rather than a non-linear regression model. The linear regression model has the advantage of being a simple statistical model that is implemented in standardised statistical packages and widely understood, whilst the direct estimation of the original non-linear form remains difficult using currently available statistical software packages (Healy 2005; StataCorp 2007; Bartholomew, Steele et al. 2008; Field 2013). It is therefore preferable to log-transform the gravity model rather than keep it in its original form.

Taking logs to linearise the mathematical gravity model is a standard approach in the literature and many researchers have followed this approach. The following will review the development of this approach in the literature. As early as 1960s, prior studies have identified that much of the misunderstanding of the gravity model is closely related to the availability of calibration methods, whilst treating the gravity model as a type of regression analysis could greatly facilitate the calibration and estimation of its parameters (Olsson 1965; Wilson 1969). Conversely, after log-transformation the gravity model could actually expand to include a whole family of spatial interaction models (Wilson 1971), which is intrinsically linear in its parameters although Equation 3.1 is in a nonlinear form itself (Batty and Mackie 1972). Therefore, linear regression analytic methods can be conveniently used to calibrate and estimate parameters of these log-transformed regression models (Draper and Smith 1966; Batty and Mackie 1972).

The calibration of parameters of linear regression models can be conventionally obtained by least squares estimation or maximum likelihood estimation, which produce the same results under normal error assumption (Draper and Smith 1966). For instance, Flowerdew (1982) used the iterative weighted least square method to calibrate the parameters of the gravity model as part of the generalised linear modelling approach. Raymer, Abel et al. (2007) used the maximum likelihood estimation to calibrate a log-linear formulation of the gravity model to estimate internal elderly migration flows. Specifically, the entropy maximization[53] is widely used in calibrating parameters of the gravity model in migration studies, as the gravity model describes spatial interaction with maximum entropy and entropy maximization is equivalent to maximum likelihood estimation (Wilson 1969; Batty and Mackie 1972; Willekens 1994; Willekens 1999). Apart from the maximum likelihood estimators (Evans 1971), the use of the entropy maximization approach in calibrating parameters of the gravity model in migration studies can be further explained by two other different theoretical justifications (Rogers and Willekens

[53] Willekens (1994) summarised the interpretation of entropy in modelling migration flows with the gravity model as follows: ‘ In the statistical interpretation of the entropy, it is found that entropy is proportional to the logarithm of the number of microscopic ways in which a given macroscopic state can be realized (the proportionality factor is the Boltzmann constant). In the case of migration, the macro-state is a given migration matrix (origin-destination table); the micro-state is an assignment of individual migrants to the origin-destination table. If information on the migration matrix is incomplete, then the most likely configuration is the one with the highest entropy, i.e. which can be produced by a maximum number of microstates that satisfy the data’ (Willekens 1994, p. 13).

1986; Willekens 1994) - information theory (Jaynes 1957), and Bayes' theorem for conditional probabilities (Hyman 1969).

Since 1980s, re-expressing the gravity model as log-linear models has become the common practice in migration studies (Flowerdew 1982; Willekens 1983; Alonso 1986; Willekens 1999; Raymer, Abel et al. 2007). In fact, the link between the gravity model and its log-linear formulation is explicit by referring the migration flow matrix as a contingency table (Alonso 1986; Willekens 1999), which further enables the application of contingency-table analysis theories and methods to migration flow data (Willekens 1999)[54]. This is a major innovation in modelling migration flows, with the underlying advantage of the unambiguous interpretation of the parameters in the log-linear formulation of the gravity model (Willekens 1994)[55].

This thesis also follows this standard approach, in order to fulfil the need to expand the selection of predictors and to allow for calibrations using mainstream statistical software packages. Based on former theoretical discussion about log-transformation, taking the natural logarithm of Equation 3.1 and adding a residual ε_{ij} results in the following log-normal linear regression formulation of the gravity model, where a, b and c are now regression coefficients to be estimated from the data

$$\ln(m_{ij}) = \ln(k) + a\ln(p_i) + b\ln(p_j) - c\ln(d_{ij}) + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \qquad (3.2).$$

[54] Willekens (1999) explicitly explained what the advances are in viewing the migration flow matrix as a contingency table 'The log-linear model and the logit model are the major tools for the analysis of contingency tables. Spatial interaction models were reformulated as log-linear models and logit models. This shift increased the interest in the application of probability theory and statistical inference in spatial interaction analysis' (Willekens 1999, p. 241).

[55] Willekens (1994) also mentioned another advantage of the linearisation of the gravity model as follows: 'The specification of spatial interaction models as log-linear models also enables the assessment of the contribution of each set of information to the predicted flow. The problem of estimating spatial interaction flows becomes a problem of including the appropriate main and interaction effects in the flow model' (Willekens 1994, p. 15).

Origin-destination distance d_{ij} is a covariate of flow-pair level, and d_{ij} is equal to d_{ji}. Residual ε_{ij} represents the deviation of the observed number of migrants from the theoretical value estimated by origin and destination population and distance, and it follows the independence assumption of linear regression. The log-log form of this model leads these parameters to be interpreted as elasticities (Leamer 2012); the relative change in the conditional expectation of m_{ij} associated with a unit relative change in the relevant covariate. For example, regression coefficient a is approximately equal to the percent increase in m_{ij} associated with a one percent increase in p_i holding all other covariates constant.

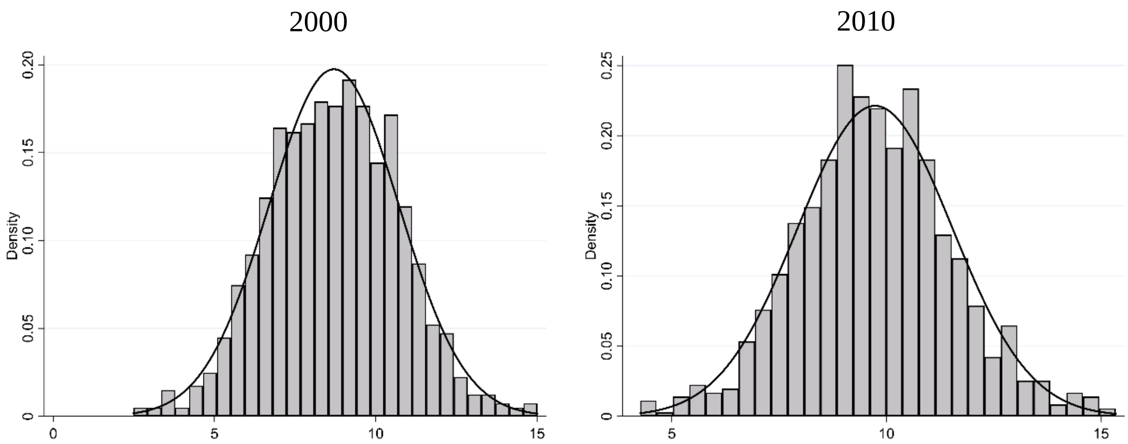

Figure 3.9 Histogram of the log transformed interprovincial migration flows in 2000 and 2010

Interprovincial migration flows are large and do not contain zero flows in 2000 and 2010, histograms of the log-transformation of interprovincial migration flows is plotted to check for the distribution of the data. Figure 3.9 provides visual evidence for that, as each histogram verifies a symmetric distribution with well-behaved tails for the log transformed interprovincial migration flows in 2000 and 2010. The histogram of log flows also closely fits the normal curve in both years (Figure 3.9). This means the log-transformed migration flow data in both years are fairly normally-distributed, symmetrical and do not have skewness issues (Figure 3.9).

3.2.3 The general form of the linear regression formula

The general form of the gravity model is a natural extension of Equation 3.2 to include further predictors of migration. It is convenient to write this more general model in vector notation as follows

$$y_{ij} = \beta_0 + \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + \mathbf{x}'_{2j}\boldsymbol{\beta}_2 + \mathbf{x}'_{3ij}\boldsymbol{\beta}_3 + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \qquad (3.3),$$

where y_{ij} denotes the log migration flows, $\mathbf{x}_{1i}$ is a vector of province-level covariates including log population and other demographic, social, environmental and/or economic attributes of origin i, $\mathbf{x}_{2j}$ is an equivalent vector of province-level covariates for destination j, $\mathbf{x}_{3ij}$ is a vector of province pair-level covariates including the log of the distance between them, and $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ are the corresponding coefficient vectors. As $\mathbf{x}_{3ij}$ only includes the log of provincial distance d_{ij} in Chapter 4 and 5, the marginal effect of log distance upon log migration is constant and equal to the one and only element in the coefficient vector $\boldsymbol{\beta}_3$.

Flowerdew and Aitkin (1982) proposed a Poisson regression version of Equation 3.3 which has various advantages over the Normal regression formulation and can become applicable when migration flows are low. The application of the gravity model in this thesis, however, consists of very large migration flows which, when log transformed, are approximately normally distributed (see Figure 3.10) and so it is the Normal rather than Poisson formulation of the gravity model that is extended in this thesis.

3.3 Discussion

This chapter has introduced the data and measurement used in this thesis to the reader. To illustrate, it has provided an overview of the data sources, the development of measurements, the data collection strategy and the data that have been collected. In particular, this chapter has discussed the definition of flow-pairs – a pair of flows of the same migration type but in opposite directions between two provinces – and the implications of this for the thesis.

Additionally, this chapter has also provided a detailed explanation of the gravity model of migration, highlighting the development of its general linear regression formulation that underpins much of the research in this thesis. This is because this formulation serves as the basis for the different model extensions presented in each of the three research chapters. Specifically, Chapter 4 will extend this basic model into a multivariate response version of the model with four simultaneously estimated equations, in order to answer the first research question 'What are the associations between China's interprovincial migration and rural and urban level of regional inequality through origin and destination population, income and distance?'. Chapter 5 will extend the standard model to a multilevel linear regression model, so as to answer the second research question 'What are the associations between China's interprovincial migration and province level of regional inequality through flow dependencies?'. Based on the multilevel linear formulation of the gravity model in Chapter 5, Chapter 6 will further extend the specification of distance decay to be a non-linear function and will simultaneously consider the contiguity of provinces and regions. This will answer the third research question 'What are the associations between China's interprovincial migration and region level of regional inequality through distance decay?'.

This chapter has shown how it is now much methodologically clearer to develop model extensions in each results chapter to answer each research question respectively. With this chapter having explained general features of data and basic model, specific data and model extensions will consequently be further addressed within each results chapter. The next chapter is the first research chapter.

Chapter 4 Analysis of the rural and urban income divide and interprovincial migration in China from 2000 to 2010 with gravity models

The previous chapter has extensively discussed the data and measurement, and has shown how the gravity model of migration has evolved in time. With this knowledge in mind, this chapter is devoted to quantifying the effect of rural and urban income in different interprovincial migration streams.

A sole-authored paper version of this chapter is published in *The Economics of Migration, Quadernidel Premio* «Giorgio Rota» N. 4, 2016 (Zhang 2016).

4.1 Introduction

Internal migration is an important topic in population studies (Otoiu 2014), and different approaches and models have been applied to China's context to quantify it (Li, Liu et al. 2014; Shen 2015). With China's rapid internal migration growth, there have been many discussions about the motivating factors underpinning the population movement (Wang, Chen et al. 2013; Liu, Xie et al. 2014). To date, there has been an emphasis on studying rural-urban migration, due to its prominent share in total migration and its significant social impacts (Ye, Wang et al. 2013).

Interprovincial migration has greatly facilitated the mobilisation of capitals, knowledge and other resources between provinces (Fan 2005a; Wang, Wei et al. 2005; Fu and Gabriel 2012). Over the last few decades in China, the volume of interprovincial migration has been growing rapidly, with interprovincial migration more than doubling from 42 to 86 million between 2000 and 2010 based on the life-long migration records in the census data[56]. It is during this period that China experienced a

[56] See Sub-section 3.1.1 about limitations using life-time migration records.

remarkable socio-economic transition and restructuring (Zhang, Lue et al. 2004; Wei 2007; Lu and Gao 2011) with interprovincial migration playing a crucial role during this process (Fan 1999; Shen, Feng et al. 2006; Bao, Bodvarsson et al. 2011; Lu and Gao 2011). For instance, rural and urban income grew by 33% and 21% respectively[57]. Meanwhile, urban population increased by 26% whilst rural population only experienced a 0.6% increase[58]. Therefore, it is important to examine interprovincial migration in 2000 and 2010 to contextualise how the migration system change during the decade[59].

There are two gaps in the literature, regarding China's interprovincial migration studies of this period. One is the lack of investigation about the heterogeneity of migration flows, as not all the migration flows happening within China originate from rural and end at urban locations. Urban-to-urban, rural-to-rural and urban-to-rural flows also occur. The other is how much income divides between rural and urban areas are related to migration flows.

To address these two knowledge gaps, this chapter adopts a new approach that specifies the gravity model with rural and urban populations and incomes. By doing so, it is possible to estimate the contribution of rural and urban incomes in China's interprovincial migration for the years 2000 and 2010, and investigate the migration flows of four different types in 2010. In the model, the interprovincial migration flow from Province A to B is considered as being composed of four streams: urban-urban, urban-rural, rural-urban and rural-rural migration. Apart from the distance between provincial capitals, urban and rural population sizes and incomes within provinces are used to model the migration flows.

[57] Measured by 1978's price. The result may be subject to minor differences from calculations using latest data realised by China's National Bureau of Statistics due to their constant adjustments of historical data. Data sources are Chinese Yearbook 2001 and 2011 (2014's release by China's National Bureau of Statistics). See Sub-section 3.1.2 for more details.
[58] Data sources are Chinese Census 2000 and 2001. See Sub-section 3.1.2 for more details.
[59] The changing context of the migration system in China is described in more general and extensive terms in Chapter 2. See Sub-section 2.3.1.1 and 2.3.1.4 for more details.

This chapter's structure is as follows. To begin with, this section subsequently provides the necessary research background. The second section conducts a review of relevant research, in order to map out the need to develop new specifications of the gravity model for this study. The third section provides a description of the data specific to this chapter, followed by a detailed explanation of the two models developed respectively for studying the total migration flow in 2000 and 2010 and the four streams of the interprovincial migration in 2010. The fifth section is devoted to presenting major results, followed by a discussion of key findings. Finally, the conclusion section suggests a future research plan for studying China's interprovincial migration.

4.2 Rural and urban income divide, internal migration and the gravity model

Rural and urban divides have been widely studied in internal migration research, and regarded as the major incentive for regional population movement, particularly rural-urban migration (Rodgers 1983; Zhu, Bell et al. 2013). It has been long believed that access to employment and income as well as social facilities such as education and health services generally are greater in urban areas and that these encourage population movement migrating from the countryside to the city (Rodgers 1983; Mohtadi 1990; Mendola 2012). Multiple empirical studies found strong evidence in both developed and developing countries for a link between rural-urban divide (notably the income divide) and rural-urban migration (Zhao 1999; Seyfrit, Bjarnason et al. 2010; Mendola 2012; Villarreal and Hamilton 2012).

The rural and urban income divide is conventionally interpreted as a one-dimensional divide from place A's rural to place B's urban area, implying that urban areas always remain favourable for internal migrants (Ravenstein 1885; Petersen 1958; Berry 1976; Todaro and Stilkind 1981). Though there is undoubtedly some truth to this interpretation, one province's rural areas, for example, are not necessarily less attractive than the other province's rural or even urban areas. The traditional interpretation, therefore, typically ignores the fact that there can also exist three other income divides, namely the divide between rural areas of place A and B, the divide between urban areas of place A and B, and the divide from place A's urban to place B's rural area.

There has been a lack of investigation into the four types of income divide (urban-urban, urban-rural, rural-urban and rural-rural) and their relationship with migration flows in China. Indeed, it is the four types of the rural and urban income divide that underpin the variations in choices of the destination and thus create migration flows of different directions in a regional system. Thus, there is a need to draw on relevant studies to fully investigate the linkages and interactions of the rural and urban areas between different places through the four types of the rural and urban income divide (Sheng 2011; Luo, Shen et al. 2014; Shen 2016b).

This chapter chooses to extend the gravity model of migration in a new way to investigate the interprovincial migration in China. This modelling approach means that a rural migrant from Province A has three types of potential places to migrate to: urban areas within Province A itself, and either rural or urban areas in another province (Province B for example). Though the focus of this study is the interprovincial movement, it is still important to consider the competition of all the three types of places in attracting this rural migrant. The flexibility of the gravity model allows for including both the rural and urban factors of Province A (the origin) and Province B (the destination) in the analysis.

Although the gravity model of migration is widely regarded as a good tool and the most popular model analysing migration flows, studies using it to directly model the relationship between the internal migration flows and the rural and urban divides in China are relatively scarce (Fan 2005b). In the few studies directly modelling China's internal migration with gravity models, total population and GDP (or GDP per capita) are used (Wang 1993; Fan 2005b; Mi, Zhou et al. 2009; Shen 2015) - these do not contain both rural and urban specific information and thus are unable to measure the role of the rural and urban incomes simultaneously in interprovincial migration. This study proposes a new way to specify the gravity model, where total population and GDP (or GDP per capita) are replaced with within-province rural and urban populations and incomes. The selection of these variables is also in line with findings in former studies, wherein the size of a place's population and the measure of its wealth have been found to be the main drivers of migration (Christian and Braden 1966; Claeson 1969; Johnston 1970; Ginsberg 1972; Fan 2005b; Shen 2012).

Another essential migration determinant of the gravity model is the distance between the origin and the destination representing geographical friction, which is measured as the distance between provincial capitals in the conventional way (Fan 2005b; Liu, Stillwell et al. 2014; Huo, Wang et al. 2016). Particularly, the impacts of the Hukou system, which is widely regarded as the root of rural-urban divide (Liu 2005) and the institutional barrier to internal migration (Zhan 2011; Chen, Liu et al. 2013; Wang, Guo et al. 2015b), could also be represented by the distance to an extent. Recalling the discussion of the Hukou system[60], it is inevitable for migrants to face with certain hindrance of the Hukou system during the change of the residential location from the origin province (Province A for instance) to the destination province (for example, Province B). Moreover, four migration streams from the origin to the destination province are confronted with different adverse forces given the same physical distance between Province A and Province B. For example, rural migrants of Province A could experience different extents of institutional difficulty in moving to rural or urban areas of Province B, when the migratory distance between Province A and Province B remains constant. Consequently, it is expected in this study to find that the effect of distance differs between the four migration streams. This chapter will also test whether these differences are statistically significant.

To summarise, this chapter takes up the challenge to examine the relationship between the interprovincial migration flows and rural and urban income divide, by using within-province rural and urban populations and incomes along with the between-province distances in specifying the gravity model. The results are then discussed in comparison to existing studies and knowledge.

[60] See Sub-section 2.3.1.2 of Chapter 2 for more details.

4.3 Data

All the variables are listed in Table 3.1, drawn from Census (2000 and 2010) and China Statistical Yearbook (2001 and 2011). Specifically, migration and population data come from the censuses in the year of 2000 and 2010[61], whilst the income data are from China Statistical Yearbook in 2001 and 2011 respectively[62]. All the independent population variables and the response variable of the total number of interprovincial migration are from the short-form dataset for the whole population contained in the censuses for both years[63]. The four streams of the interprovincial migration come from the long-form of the 2010 Census. It is worth emphasizing that data for the four streams are only available in 2010, which sets further limits in selecting predictors (it is not practical to use 2000's migration data to predict 2010's streams in this case) and analysis methods. Recalling that all the independent population variables exclude counts of the migrants (i.e., the dependent variable), in order to make sure the regression equations do not include the same quantity on both sides. Each variable is log-transformed in the models but is shown in its original scale in Table 4.1.

Table 4.1 Variable list[64]

Dependent variable	unit	Independent variable	unit
interprovincial migration	000s persons	origin-destination distance	000s m
urban-rural migration	00s persons	origin urban population	000,000s persons
rural-urban migration	00s persons	origin rural population	000,000s persons
rural-rural migration	00s persons	origin urban income	000s yuan
urban-urban migration	00s persons	origin rural income	000s yuan
		destination urban population	000,000s persons
		destination rural population	000,000s persons
		destination urban income	000s yuan
		destination rural income	000s yuan

61 The migration flow here is enumerated as lifetime flow counts. The reason to adopt such a definition and the detailed migration data acquisition process is presented in Sub-section 3.1 of Chapter 3.
62 The detailed data acquisition process is in Sub-section 3.1 of Chapter 3.
63 Sub-section 3.1 of Chapter 3 explicitly illustrates the specific data structure of Census 2000 and 2010.
64 Section 3.1 of Chapter 3 offers explicit explanations about measurements of origin-destination distance and incomes.

4.4 Methods

This section starts by presenting a general form of the gravity model in a linear regression formulation. Then, a univariate formulation of the gravity model is developed for total interprovincial migration in 2000 and 2010. Finally, an improved multivariate formulation is presented to analyse the four types of interprovincial migration flows in 2010.

4.4.1 A univariate formulation of the gravity model - total interprovincial migration in 2000 and 2010

The linear regression formula of the general form[65] (Equation 3.3) can be written as

$$y_{ij} = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{x}_j'\boldsymbol{\beta}_2 + \mathbf{x}_{ij}'\boldsymbol{\beta}_3 + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \qquad (4.1)$$

where $\mathbf{x}_i'$ represents a vector of factors measuring the demographic, social, environmental and/or economic attributes of origin i, $\mathbf{x}_j'$ represents a vector of factors for destination j, $\mathbf{x}_{ij}'$ denotes the distance between *i* and *j*, and $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ are the corresponding coefficient vectors for $\mathbf{x}_i'$, $\mathbf{x}_j'$ and $\mathbf{x}_{ij}'$.

Equation (4.1) maybe further specified as $\mathbf{x}_i' = [\ln(p_{ui}) \ln(p_{ri}) \ln(I_{ui}) \ln(I_{ri})]$ and $\mathbf{x}_j' = [\ln(p_{uj}) \ln(p_{rj}) \ln(I_{uj}) \ln(I_{rj})]$ in this study, where p_{ri} and p_{ui} represent the rural and urban population counts for each province and I_{ui} and I_{ri} the urban and rural incomes of the origin i. $\mathbf{x}_j'$ are

[65] See Sub-section 3.2.3 of Chapter 3 for more details.

the corresponding values for the destinations. Including the incomes reflects the importance of economic factors in the migration system (Fan 2005b; Beine, Bertoli et al. 2014).

4.4.2 A multivariate formulation - four types of interprovincial migration flows in 2010

Whilst this univariate response linear regression is adequate to analyse the total interprovincial migration in 2000 and 2010, a multivariate response linear regression approach is preferable to conduct the analysis of the four streams in 2010. This is because it is not easy to statistically compare and test model coefficients across linear regression models when they are estimated separately and their residuals are assumed independent of the responses and also un-correlated with each other. In contrast, multivariate response linear regression can be used to estimate the four equations jointly, whilst the residuals are assumed to be not only independent between (random) observations but (possibly) correlated between/among the responses for each observation. These advantages could therefore facilitate post-estimation model comparisons and tests (Holm 1979; Simes 1986). To illustrate, such post-estimation model tests[66] of a single joint model take into account of the correlation among the responses (and therefore their associated residuals of each equation respectively), which effectively enables statically comparing and testing coefficients across the four equations in a more precision manner.

In developing a multivariate formulation of the gravity model, Equation (4.1) is extended to include four responses y_{1ij}, y_{2ij}, y_{3ij} and y_{4ij}, using the same set of predictors $\mathbf{x}_i'$, $\mathbf{x}_j'$ and $\mathbf{x}_{ij}'$ but with different coefficients as shown below:

[66] Post-estimation tests can be univariate or high-dimensional. Univariate post-estimation test is easy to conduct and can provide straightforward test results regarding a chosen variable, but it might also falsely reject one of the null hypotheses when testing multiple variables simultaneously. High-dimensional post-estimation test is therefore developed to account for this issue (Holm 1979; Simes 1986). In this chapter, univariate post-estimation tests are first conducted, followed by high-dimensional post-estimation tests to bind the probability of falsely rejecting one of the null hypotheses. By doing so, it is possible to test individual variables separately and all the four stream equations simultaneously as a whole.

$$y_{nij} = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta}_{n1} + \mathbf{x}_j'\boldsymbol{\beta}_{n2} + \mathbf{x}_{ij}'\boldsymbol{\beta}_{n3} + \varepsilon_{nij}$$

$$\begin{pmatrix}\varepsilon_{1ij}\\ \varepsilon_{2ij}\\ \varepsilon_{3ij}\\ \varepsilon_{4ij}\end{pmatrix} \sim N\left\{\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}\sigma_{\varepsilon 1}^2 & & & \\ \sigma_{\varepsilon 21} & \sigma_{\varepsilon 2}^2 & & \\ \sigma_{\varepsilon 31} & \sigma_{\varepsilon 32} & \sigma_{\varepsilon 3}^2 & \\ \sigma_{\varepsilon 41} & \sigma_{\varepsilon 42} & \sigma_{\varepsilon 43} & \sigma_{\varepsilon 4}^2\end{pmatrix}\right\}$$

$$n = (1, 2, 3, 4) \qquad (4.2).$$

In Equation (4.2), y_{1ij}, y_{2ij}, y_{3ij} and y_{4ij} each stand for the streams of urban-urban, urban-rural, rural-urban and rural-rural interprovincial migration. ε_{1ij}, ε_{2ij}, ε_{3ij} and ε_{4ij} denote the variances for the four responses respectively. It has to be pointed out that multivariate regression generates the same covariate coefficients with equation-by-equation univariate regression for the individual responses. However, fitting separate equations will allow for different variances. Nevertheless, the residuals associated with the four different responses may have correlated variances, whose correlations can be derived from the associated variance and covariance parameters in the usual way (for instance, $\rho_{\varepsilon 11} = \sigma_{\varepsilon 11}/\sigma_{\varepsilon 1}^2$ and $\rho_{\varepsilon 21} = \sigma_{\varepsilon 21}/\sigma_{\varepsilon 1}\sigma_{\varepsilon 2}$). Such correlation between the four residuals will be neglected if estimated separately by the univariate regression equation. Multivariate regression, in this regard, enables the further investigation of the correlation and association of the residuals, which will offer new insights into the four streams in 2010. In other words, procedures for statistical inference in the multivariate linear regression take account of the fact that these four streams can be correlated responses[67] (Fox and Weisberg 2011).

[67] There is indeed strong correlation among the four responses (correlation coefficients 0.80~0.92, *p*<0.001).

4.5 Results

This section consists of two sub-sections. Results of 2000 and 2010 total migration are explained in the first sub-section, whilst the second sub-section details results of 2010's four interprovincial migration streams.

4.5.1 2000 and 2010 total migration

To begin with, Figure 4.1 presents an overall picture of the primary spatial pattern for interprovincial migration, detailing the 30 largest out of a total of 930 interprovincial migration flows in 2000 and 2010. Specifically, the 30 largest interprovincial migration flows accounted for 57.37% and 51.97% of the total interprovincial migration in 2000 and 2010 respectively, and they are classified into three groups by tertiles within each year: 'Small', 'Medium' and 'Large'. The reason behind this is to observe and compare the major spatial patterns in both years. Each group has 10 flows for both years, but the absolute number varies notably for the same group between the two years. For instance, 10 flows ranging from 232 thousand to 323 thousand in 2000 and 10 flows ranging from 570 thousand to 899 thousand in 2010 are both categorised as 'Small'. 'Medium' contains 10 flows ranging from 335 thousand to 750 thousand in 2000 and 10 flows ranging from 934 thousand to 1,530 thousand in 2010, whilst 'Large' contains 10 flows ranging from 782 thousand to 3,329 thousand in 2000 and 10 flows ranging from 1,560 thousand to 4,602 thousand in 2010. The rest of the 812 flows for each year, by comparison, are small in volume and could not represent the prime spatial features of interprovincial flows. Therefore, the rest of the 812 flows are not shown in the figure[68].

[68] Both the correlation and rank correlation stand at 0.96 (p<0.001) between total migration in 2000 and 2010. See Subsection 3.1.1 for more detailed comparison of total migration in 2000 and 2010. The histogram of total migration in 2000 and 2010 is presented in Sub-section 3.2.2.

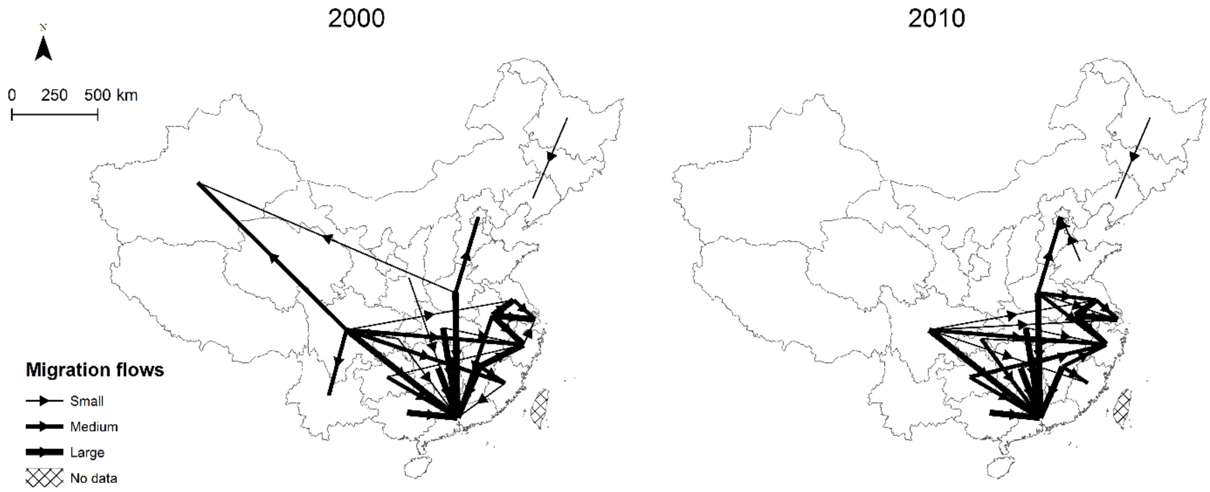

Figure 4.1 The 30 largest interprovincial migration flows in 2000 and 2010

For both years, Figure 4.1 reveals Sichuan, Anhui and Henan as leading origins, and Guangdong, Zhejiang and Shanghai as the predominant destinations. Specifically, the three leading origins are from either the West or Central regions, whilst all the three leading destinations are located in Eastern China hereby referred to as the East region. Notably, none of the top 30 leading interprovincial migration destinations are in the Central region in 2000 or 2010. This reflects the predominant role of the Central region as the source of most migrants and highlights the uneven regional development in China, whereby the East region is the most developed followed by the Central region and West region respectively. Additionally, Xinjiang and Yunnan lost their prominence in attracting in-migrants in 2010, losing the only two destinations that appeared from West China in the Top-30 leading interprovincial migration flows in 2000. This may imply that provinces in the West were further disadvantaged in the regional migration system between 2000 and 2010. Meanwhile, migration to Beijing and Shanghai underwent notable growth, whilst Henan from the Central region experienced a marked rise in out-migration, indicating the growing linkages between the Central and East regions during this period. Between 2000 and 2010, another noticeable change was in the migratory distance, with larger flows observed taking place between adjacent provinces.

Table 4.2 The total interprovincial migration model results

	2000		2010	
Parameter	Estimate	SE	Estimate	SE
Constant	9.958***	0.912	7.453***	1.052
Log of origin urban population	0.166	0.113	0.480***	0.091
Log of origin rural population	0.943***	0.083	0.669***	0.077
Log of destination urban population	0.636***	0.113	0.942***	0.091
Log of destination rural population	0.017	0.083	-0.330***	0.077
Log of origin urban income	-1.118***	0.322	-1.715***	0.377
Log of origin rural income	0.492	0.272	0.104	0.289
Log of destination urban income	2.002***	0.322	4.327***	0.377
Log of destination rural income	0.155	0.272	-1.437***	0.289
Log of distance	-1.096***	0.067	-1.002***	0.053

Note: Response variable is the log migration flow (in 1000 persons). * denotes p < 0.05; ** denotes p < 0.01; *** denotes p < 0.001.

Table 4.2 shows the results for the total interprovincial migration in 2000 and 2010, using the univariate linear regression model (Equation 4.1). Both models include the natural log of the origin and the destination rural and urban populations and incomes as covariates as well as the natural log of the distance between each pair of provinces. The adjusted R^2 are 0.653 ($F = 195.57, p < 0.001$) and 0.737 ($F = 290.03, p < 0.001$) for 2000 and 2010 respectively, indicating that the covariates fit well with the observations of total interprovincial migration in both models.

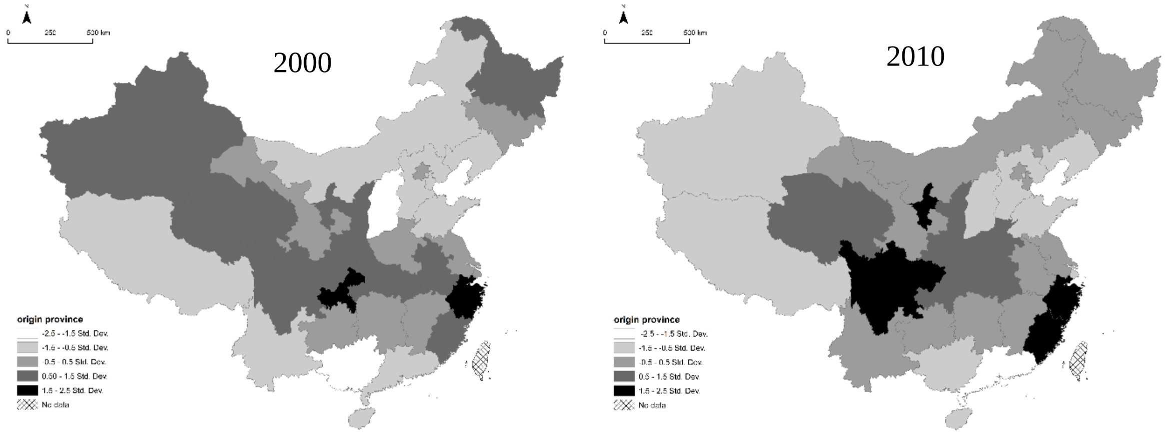

Figure 4.2 Model standardised origin-level residuals in 2000 and 2010

Figure 4.2 maps the standardised origin-level residuals (i.e., the average flow residual in each origin province expressed in SD units) in 2000 and 2010, showing that the standardised residuals of most provinces fall within the range of -1.5 and 1.5 standard deviation[69] in both years. In 2000, Guangxi and Shanxi have standardised average residuals ranging from -2.5 to -1.5, meaning that flows originating from Guangxi and Shanxi tend to be overestimated by 2000's model. Guangxi had a very low urbanisation rate (17.72%) and poorly-developed infrastructure in 2000 (Shenggen and Zhang 2004), which may have contributed to its lower-than-expected emissivity. By comparison, Shanxi had an urbanisation rate (25.53%)[70] higher than the national average (24.73%)[70] and a strong mining industry (Lee 2003), which may have had facilitated the distribution of migrants within the province and thus led to the lower-than-expected interprovincial out-migration flows in 2000. Chongqing and Zhejiang have standardised residuals ranging between 1.5 and 2.5 standard deviation, meaning that flows originating from both provinces tend to be underestimated by 2000's model. Chongqing and Zhejiang both had a very low share of primary sector respectively (17.8% and 11.0%)[71] but a relatively high share

[69] Here, residuals outside the range of -1.5 and 1.5 are considered to represent poor fit. That is, average residuals exceeding the range of -1.5 and 1.5 are treated as showing un-modelled between-province heterogeneity in this case. This between-province heterogeneity is effectively because the model assumes no clustering. A common criteria in the literature is the range of -2 and 2. But the residuals here are actually average residuals, so I adopt a more conservative criteria (Serra, 2002).
[70] Data source is China Census 2000.
[71] Data source is China Yearbook 2001.

of agricultural population (65.41% and 78.73%)[70], which may have led to the higher-than-expected out-flows of interprovincial migration. In 2010, Guangdong had standardised average residuals ranging from -2.5 to -1.5, whilst Ningxia, Sichuan, Zhejiang and Fujian have standardised average residuals ranging between 1.5 and 2.5. Interprovincial flows originating from Guangdong tend to be overestimated by 2010's model, which may be caused by its huge total population [72](103 million, 1st of provincial population in China)[73]. Interprovincial flows originating from Ningxia, Sichuan, Zhejiang and Fujian are underestimated by 2010's model possibly for two different reasons: they all had low shares of primary sector (4.9% ~ 14.4%)[74] but relatively high shares of agricultural population (78.36% and 59.82%)[73]. To summarise, factors such as provincial urbanisation rate (i.e. the share of provincial urban population in the total provincial population) and industry (or economy) structure might be helpful to further refine the current model results by potentially reducing the standardised origin-level residuals. Whilst this thesis does not pursue expanding the current selection of independent variables, future research is encouraged to search for introducing additional explanatory variables into the model[75].

[72] Although the model has adjusted rural and urban population specifically, the size of the total population (the sum of rural and urban population) may still be effective in influencing interprovincial migration. This may also point to a potential future research direction, which enters population into the model in a more flexible manner.
[73] Data source is China Census 2010.
[74] Data source is China Yearbook 2011.
[75] See Section 4.7 for more information about future research directions.

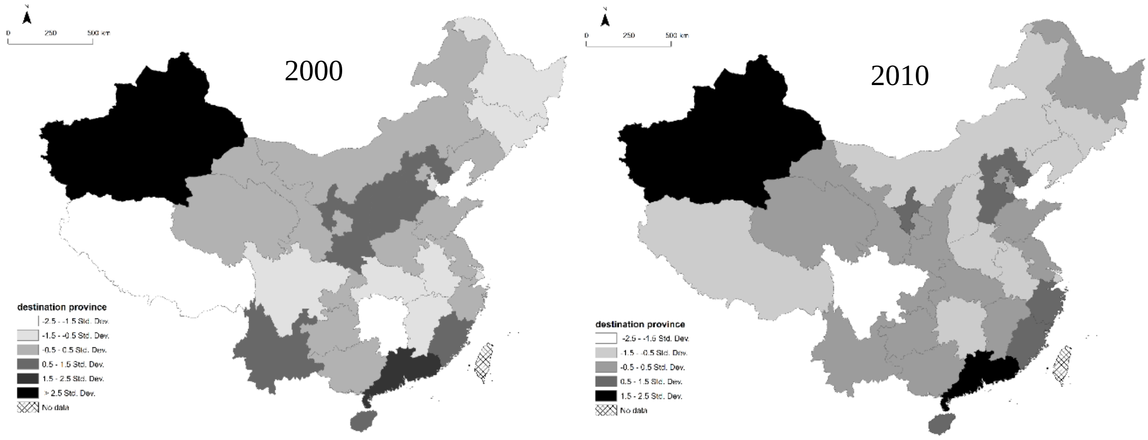

Figure 4.3 Model standardised destination-level residuals in 2000 and 2010

Figure 4.3 shows the standardised destination-level residuals (i.e., the average flow residual in each destination province expressed in SD units) in 2000 and 2010. Although standardised average residuals of most provinces fall within the range of -1.5 and 1.5 standard deviation[69] in both years, there are provinces showing poor model fit in each year. In 2000, Tibet and Guizhou have standardised average residuals ranging from -2.5 to -1.5 standard deviation, meaning that flows ending at Tibet and Guizhou tend to be overestimated by the model. This might be because both provinces had high shares of the primary sector (30.9% and 27.3%)[71] and poorly-developed infrastructure in 2000 (Shenggen and Zhang 2004). By comparison, Xinjiang attracted higher-than-expected interprovincial in-migration flows in 2000 (the standardised average residual is larger than 2.5 standard deviation), reflecting the state policy in encouraging immigration into Xinjiang (Becquelin 2000; Howell and Fan 2011). In 2010, Sichuan has standardised average residuals ranging from -2.5 to -1.5 standard deviation, whilst Xinjiang and Guangdong have standardised average residuals ranging between 1.5 and 2.5 standard deviation. Interprovincial flows ending at Sichuan tend to be overestimated by 2010's model, which may be caused by its huge total population[72] (80 million, 4th of provincial population in China)[73] and a relatively large share of the secondary and tertiary sectors (85.6%)[74]. Interprovincial flows ending at Xinjiang and Guangdong are underestimated by 2010's model possibly for two different reasons respectively: the state policy in encouraging immigration into Xinjiang (Becquelin 2000; Howell and Fan 2011), and the huge total population (103 million, 1st of provincial population in China)[73] and the large share of secondary and tertiary sectors (95%, 1st among all provinces in China)[74] in Guangdong. Similar to the earlier interpretation of the standardised origin-level residuals, factors such as state policies and industry (or economy) structure might be helpful to further improve the current model results by potentially reducing the standardised destination-level residuals.

However, the interpretation of the standardised origin-level and destination-level residuals is based on the model ignoring clustering or group effects. That is, these origin-level and destination-level residuals are the average flow residual in each origin and destination province respectively, so the origin and destination effects are in fact ignored by the model. It runs the risk of spuriously precise coefficients, as the province differences have not been formally studied. The multilevel approach will allow for all of this, which motivates the research direction of my next chapter[75].

In both models, most estimated coefficients are in the expected directions and statistically significant at a 99% confidence level. The models show that the larger the rural or urban population of the origin, the greater the flow. Similarly, the urban population size of the destination is significantly positively associated with the total interprovincial migration. By comparison, the rural population of the destination and the total interprovincial migration are less strongly related in 2000, and even significantly negatively associated with each other in 2010. The log of urban income is significantly negative in the origin but strongly positive in the destination, whilst rural income mostly remains insignificant except for the significant negative association with the destination in 2010. As expected, the distance between the provinces acts as a significant impeding factor, meaning the further the distance the weaker the flow.

In 2000, rural population and urban income of the origin, and urban population and urban income of the destination were particularly strong predictors for the direction and magnitude of migration flows, in addition to distance. Specifically, a 10% increase in the origin rural population is associated with a

9% increase in out-migration, all else held equal, while a 10% increase in the destination urban population is associated with a 6% increase in in-migration. As expected, an increase in the origin income could exert some competing attraction to the migrants and decrease the interprovincial population movement. For instance, a 10% increase in the origin urban income is associated with an 11% decrease in out-migration while a 10% increase in destination urban income is associated with a 20% increase in in-migration. In terms of the distance, a 10% decrease is associated with 11% increase in the migration flow.

In 2010, all the covariates were statistically significant predictors, except for the origin rural income. A 10% increase in the destination rural population is associated with a 3% decrease in in-migration, controlling all the other covariates. This negative association is significant and surprising as it is seldom observed in other migration studies. In fact, the destination population tends to be found to be positively related to migration flows in former studies (Piotrowski and Tong 2013; Gong and Huybers 2015; Thomas, Stillwell et al. 2015), as population of the destination usually represents the market size and employment opportunities and thus creates the fundamental attraction towards in-migrants (Clark 1982; Greenwood 1985; King 2011). For this reason, it is surprising to discover that destination rural population in this study stands as a deterring factor, though it is possible to find a positive association between destination rural population and in-migration to become weak as other studies have shown that the bigger a place's rural population is the less attractive it is to migrants (Poveda 2007; Huang, Lu et al. 2011; Ebenstein and Zhao 2015). Under this circumstance, one possible cause may lie in the systematic preference among migrants, wherein urban Hukou is generally more desirable and in-migrants from other provinces would be more willing to make the movement if the destination has a smaller rural population and thus is less likely to provide in-migrants with the rural Hukou. Similarly, the result also shows another noticeable negative association between the destination rural income and the interprovincial migration; a 10% increase in the former is significantly associated with a 14% decrease in the latter. Again, this significant negative association is counter-intuitive and seldom observed in former studies. In fact, migration theories and empirical studies have repeatedly pointed out that the growth of the destination income incurs the rise of in-migration (Greenwood 1985; De Haas 2007a; King 2011). Regarding this case, a possible explanation might be that an increase in destination rural income does not necessarily translate into the growth of the (desirable) employment opportunities in destination provinces, as rural jobs are usually offered by low-paid and labour-intensive industries which may not exert much attraction to in-migrants (Lewis 1954a; Zhang and Song 2003; Luo, Gao et al. 2011; Lemoine, Poncet et al. 2015). Thus, both negative associations demonstrate that the destination is connected to the origin through migration flows in a much more complicated way in that some factors in the destination could actually discourage in-flows and their effects could be significant and sizeable.

Between 2000 and 2010, the total interprovincial migrants rose from 42 million to 86 million, signposting a dramatic change in the between-province migration system. Although directly comparing the models between 2000 and 2010 remains difficult due to both models being estimated by two separate univariate linear equations[76], it is still possible to catch a glimpse of the system through changes in covariates within the 10 years. To begin with, provincial capital distances changed little and thus are held constant in the models, and rural population remained almost the same with a slight 0.6% national increase. That is to say, the reason that interprovincial migrants more than doubled between 2000 and 2010 lies in the changes in other covariates, namely the (rural and urban) income and the urban population (or other unmeasured factors). In fact, rural and urban income each encountered a 33% and 21%[57] growth respectively, whilst urban population experienced a 26% increase. However, it is still not clear what changes took place in the interprovincial system that made 2010's interprovincial migration so distinctly different from that of 2000. Therefore, a much closer investigation into 2010's interprovincial migration will be carried out in the next sub-section.

[76] See Sub-section 4.5.1 for more details. Univariate linear equations do not assume either correlations between different response covariates or correlations between different residuals.

4.5.2 2010's four interprovincial migration streams

In this sub-section, 2010's four interprovincial migration streams, namely urban-urban, urban-rural, rural-urban and rural-rural, are described with the major 30 flows first, and then modelled with multivariate linear regression using the same set of predictors as the last sub-section. After the modelling, some testing of the predictors and residuals are carried out to provide a more detailed investigation of the four streams.

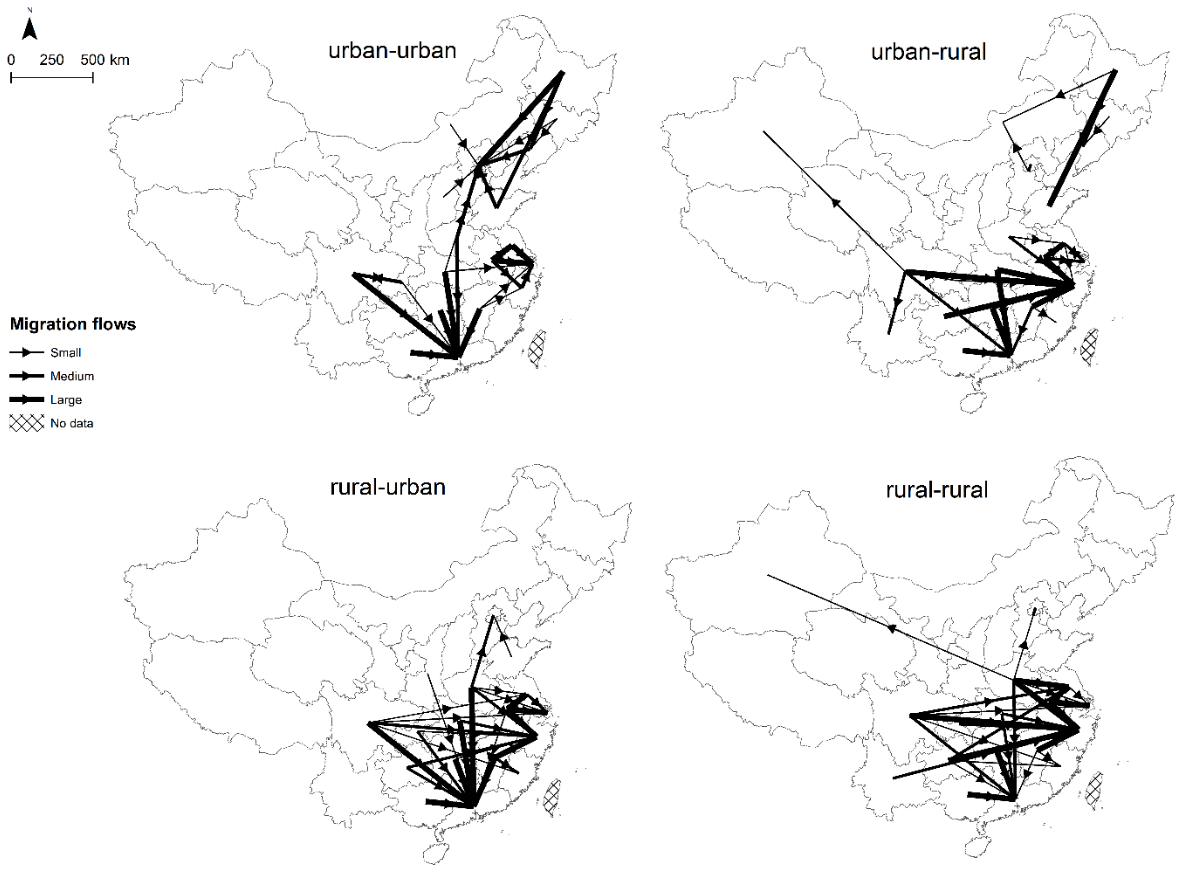

Figure 4.4 The 30 largest interprovincial migration streams in 2010

In Figure 4.4, prime spatial patterns for the four interprovincial migration streams are illustrated, depicting the 30 largest flows for each migration type in 2010. Similar to what was explained in the previous sub-section, the remaining 812 flows for each migration type are not shown in the figure in order to better outline the major flows. The 30 largest interprovincial migration flows accounted for 37.8%, 30.3%, 59.8% and 57.6% of the urban-urban, urban-rural, rural-urban and rural-rural interprovincial migration respectively. Within each migration type, the 30 largest flows are classified into ‘Small’, ‘Medium’ and ‘Large’ three groups by tertiles to compare their spatial patterns within and across different migration types.

For all the four streams, Figure 4.4 reveals Guangdong, Zhejiang, Jiangsu and Shanghai as the leading destinations and Sichuan, Henan, Hubei and Anhui as main origins. All the four leading destinations are in the Eastern region, whilst all the four leading origins are from the Central region except for Sichuan located in the West. This re-confirms the earlier finding about the dominant role of the Central region in exporting migrants and the uneven regional development in China, whilst the consistent role of Sichuan in exporting migrants may stem from its large population as well as its relatively adjacent location to the popular destinations. Compared with the other three migration types, large rural-urban migration flows tend to take place between adjacent places, among which Guangdong is the most popular. For urban-urban streams, Beijing is a remarkably popular destination and most in-flows are from neighbouring provinces and provinces of the Northeast. This indicates that Beijing is well-connected with the urban areas of the surrounding and the Northeast provinces, and that it is the centre of the north cluster in the urban-urban migration sub-plot. Xinjiang, by contrast, maintains as the only leading destination in the West for both urban-rural and rural-rural streams. Long-distance moves from urban Sichuan to rural Xinjiang as well as those from rural Henan to rural Xinjiang, may reflect the long-standing state policy in encouraging emigration to Xinjiang to develop specific agricultural industries such as growing cotton (Becquelin 2000; Howell and Fan 2011).

Table 4.3 shows the results for the four streams, and the adjusted R^2 are 0.715 ($F = 251.05, p < 0.001$), 0.561 ($F = 128.26, p < 0.001$), 0.707 ($F = 241.85, p < 0.001$) and 0.630 ($F = 170.55$, $p < 0.001$) for urban-urban, urban-rural, rural-urban and rural-rural migration respectively. The results indicate that the four equations are well fitted overall. Nevertheless, urban-rural migration seems the least predictable by the chosen set of predictors, which may indicate that it is less economically driven compared with other streams. Furthermore, correlation coefficients[77] between the four residuals are all significant (p <0.001), ranging from 0.69 to 0.83.

Table 4.3 The multivariate interprovincial migration model results

	Urban-urban		Urban-rural		Rural-urban		Rural-rural	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Constant	5.586***	0.869	6.749***	1.145	6.049***	1.268	5.395***	1.466
Log of origin urban population	0.635***	0.079	0.804***	0.104	-0.071	0.116	0.316*	0.134
Log of origin rural population	0.160*	0.064	0.120	0.085	1.361***	0.094	0.954***	0.108
Log of destination urban population	1.037***	0.079	0.249*	0.104	0.687***	0.115	0.586***	0.133
Log of destination rural population	-0.438***	0.064	0.213*	0.084	-0.219*	0.093	0.036	0.108
Log of origin urban income	-2.258***	0.313	-1.491***	0.413	-3.535***	0.457	-0.951	0.528
Log of origin rural income	1.119***	0.240	-0.067	0.316	1.012**	0.350	-1.046**	0.404
Log of destination urban income	3.738***	0.312	2.417***	0.410	4.455***	0.455	3.457***	0.525
Log of destination rural income	-1.420***	0.240	-0.890**	0.316	-1.084**	0.350	-0.938*	0.405
Log of distance	-0.854***	0.043	-1.045***	0.057	-0.991***	0.063	-1.302***	0.073

Note: Response variable is the log migration flow (in 1000s). * denotes p < 0.05; ** denotes p < 0.01; *** denotes p < 0.001.

In all the four equations, most estimated coefficients are statistically significant across various levels and in the expected directions (Table 4.3). Overall, the equations show that the larger the destination urban population or income, the greater all the streams. By contrast, it is also found that the larger the destination rural income or the greater the distance, the weaker all the streams. Moreover, origin urban population shows significantly positive association with all streams except for rural-urban migration, wherein the association is negative but insignificant. Similarly, origin rural population is also significantly positively associated with all streams except for urban-rural migration, where the association becomes insignificant but still remains positive. In addition, a significantly negative

[77] As correlation between the residuals is not the major focus here and only relevant to the post-model tests, Table 4.3 therefore does not specifically report the correlation coefficients. The discussion of residual correlations is presented at the end of this sub-section as part of post-model tests.

association with origin urban income is constantly observed in all streams except for rural-rural migration, wherein the association loses its significance. In particular, coefficients of distance show notable variation across all streams. Specifically, a 10% increase in distance between province capitals is associated with a 9%, 10%, 10% and 13% decrease in urban-urban, urban-rural, rural-urban and rural-rural migration respectively. This implies that the same interprovincial distance may hold the strongest and weakest friction towards rural-rural and urban-urban migration respectively, whilst urban-rural and rural-urban migration may endure similar and moderate distance friction. Distance friction here refers to the amount of resources, energy, cost, and time to overcome the origin-destination distance to mobilise migrants from the origin and relocate them to the destination (Williamson 1981; Eldridge and Jones III 1991). That is to say, given the same between-province migratory distance, rural-rural migration takes the largest amount of resources, energy, cost, or time to overcome and rural-rural migrants are thus most sensitive to it, whilst urban-urban migration constitutes the least amount of distance friction and urban-urban migrants are the least sensitive to it.

Destination population and origin income of the rural areas show the most variation in terms of both significance levels and effect signs. For instance, though destination rural population stays significantly negatively associated with both urban-urban and rural-urban migration, the association with urban-rural and rural-rural streams is positive but insignificant for the latter. Similarly, origin rural income is significantly positively associated with both urban-urban and rural-urban migration, but negatively associated with both rural-rural and urban-rural streams. Moreover, its negative association with urban-rural migration is not significant.

In the urban-urban stream equation, all the covariates are significant predictors. Specifically, the effects of the destination populations are worthy of attention. The destination urban population stands out due to its large effect size, with a 10% increase in it is associated with a 10% increase in urban-urban in-migration. This positive association is in line with other studies: growth of destination urban population size increases market size and employment opportunities, which is attractive for the in-migrants

(Claeson 1969; Lu and Wang 2013; Shi, Zheng et al. 2014). By contrast, the destination rural population is unusual because of its negative effect sign, indicating its negative association with urban-urban migration. This is perhaps due to the fact that growth of the destination rural population will result in increase of market size and employment opportunities for the destination province's rural areas, thus creating a diverting attraction for the urban migrants from other provinces; as a 10% increase in the destination rural population is associated with a 2% increase in urban-rural migration whilst maintaining its association with the 4% decrease in urban-urban migration. Regarding urban income, its effect is sizeable but in contradictory directions in the origin and the destination; its 10% decrease in the former is associated with 23% increase in urban-urban migration, whilst its 10% increase in the destination is associated with 37% increase in urban-urban streams. This finding echoes earlier studies, wherein rise of the origin's income has been proven to be effectively reducing out-migration whilst growth of the destination's income promotes in-migration (He and Pooler 2002; Deshingkar 2006; Gries, Kraft et al. 2015). Meanwhile, an opposite pattern is observed in rural income, in the sense that its effect is positive in the origin but negative in the destination. More specifically, 10% increase in the origin rural income is associated with a statistically significant 11% growth in urban-urban migration due to the rise of the origin province's overall emissivity, whilst 10% increase of the destination rural income is associated with a 14% decrease in urban-urban migration because of the diverting attraction growth from the destination rural areas. In terms of the distance, a 10% decrease in it is associated with a 9% increase in urban-urban migration.

For the urban-rural stream equation, origin urban population and destination urban income are particularly strong predictors, compared with other population and income factors respectively. Specifically, a 10% increase in origin urban population is associated with 8% increase in urban-rural migration, in line with findings of the origin population size from other studies (Bhat 2002; Liu, Dong et al. 2009; Wang, Zhang et al. 2013). Meanwhile, a 10% increase in destination urban income is associated with 24% increase in urban-rural migration, indicating that the rise of destination urban income contributes to the overall increase of the destination province's attraction towards urban migrants from other provinces. As expected, all the other population factors are positively associated with urban-rural migration, reflecting the general positive effect of populations observed in former studies (Relethford 1986; Peeters 2012; Ye, Wang et al. 2013). However, both origin urban income and destination rural income are negatively associated with urban-rural migration, re-confirming the earlier observation about the overall model fit that urban-rural migration is not as economically driven as other streams. The former negative association is also observed in urban-urban migration and can be interpreted in the same way, whilst the latter negative association may be related to the general inefficient translation from income growth to employment increase for rural areas as mentioned earlier. As with the distance, its 10% increase is associated with a 10% decrease in urban-rural migration.

As with the rural-urban stream equation, rural population and urban income of the origin as well as destination urban income are predictors worthy of attention due to the magnitude of their effect on migration in comparison with other population and income factors respectively. For the origin rural population, its 10% increase is associated with a 14% increase in rural-urban migration, echoing the positive relationship between the origin population and the out-migration observed in the general literature. For the origin urban income, its 10% increase is associated with a 35% decrease in rural-urban migration, which lies in the diverting force from the origin urban income; to be more specific, the rise of the urban income in the origin province promotes the attraction of the origin urban areas and diverts rural migrants to make within-province rather than between-province rural-urban movements (Wu and Yao 2003; Alm and Winters 2009; Brunarska, Nestorowicz et al. 2014). For the destination urban income, a 10% increase is associated with a 45% increase in rural-urban migration. This positive association stems from increase of the attraction in urban areas of the destination province accompanying the urban income growth, whilst a 10% increase in distance is associated with a 10% decrease in rural-urban migration.

Regarding the rural-rural stream equation, origin rural population and destination urban income have large effect sizes compared with other income and population factors, and consequently are predictors worthy of attention. As with origin rural population, a 10% increase is associated with a 10% increase in rural-rural migration, proving the positive association between origin population and out-migration once again. By contrast, effect sizes of all the other population factors are much smaller or even insignificant, despite also being positively associated with rural-rural migration. In terms of the destination urban income, a 10% increase is associated with a 35% increase in rural-rural migration, signifying a positive association between urban income and attraction towards in-migrants in the destination as observed in other three migration streams; in comparison, the effects of other income factors remain negative and much smaller in the effect size. For the distance, a 10% increase is associated with a 13% decrease in rural-rural migration.

Table 4.4 The comparison of standardised coefficients

Parameter	Urban-urban Standardised Coefficients	Urban-rural Standardised Coefficients	Rural-urban Standardised Coefficients	Rural-rural Standardised Coefficients
Log of origin urban population	0.366***	0.437***	-0.028	0.123*
Log of origin rural population	0.101*	0.072	0.598***	0.408***
Log of destination urban population	0.599***	0.136*	0.276***	0.229***
Log of destination rural population	-0.278***	0.127*	-0.097*	0.015
Log of origin urban income	-0.362***	-0.225***	-0.394***	-0.103
Log of origin rural income	0.281***	-0.016	0.176**	-0.177**
Log of destination urban income	0.609***	0.371***	0.504***	0.381***
Log of destination rural income	-0.361***	-0.213**	-0.191**	-0.161*
Log of distance	-0.375***	-0.432***	-0.302***	-0.386***

Note: * denotes $p < 0.05$; ** denotes $p < 0.01$; *** denotes $p < 0.001$.The standardised regression coefficients measure the standard deviation change in each dependent variable expected with a one standard deviation change in each explanatory variable.

A comparison of the predictors' effect sizes is conducted through their standardised coefficients for the four stream equations (Table 4.4), following the typical procedure of comparing coefficients across models (Schielzeth 2010). The standardised regression coefficients are simply the regression coefficients from models where with both the independent variables (X) and the dependent variables (Y) in are expressed in standard deviation units (i.e., transformed to z-scores), in order to facilitate the comparison of effect sizes of predictors within and across different equations. The values listed give the standard deviation change in Y expected with a one standard deviation change in X.

Among the standardised coefficients of population predictors, most are positive. For instance, rural population carries the most sizeable push to rural-urban stream in the origin (0.598; a one SD increase in origin rural population is associated with a 0.598 SD increase in individuals migrating from rural to urban areas between provinces) and the strongest attraction for urban-rural migration in the destination (0.127; a one SD increase in destination rural population is associated with a 0.127 SD increase in individuals migrating from urban to rural areas between provinces), whilst the negative association of destination rural population with both urban-urban (-0.278; a one SD increase in destination rural population is associated with a 0.278 SD decrease in individuals migrating from urban to urban areas between provinces) and rural-urban streams (-0.097; a one SD increase in destination rural population is associated with a 0.097 SD decrease in individuals migrating from rural to urban areas between provinces) signposts the diverting effect of rural population in the destination provinces. However, this diverting effect does not happen to destination urban population, as its coefficients remain positive for all the streams (Table 4.4). More specifically, destination urban population has the largest association with urban-urban stream (0.599; a one SD increase in destination urban population is associated with a 0.599 SD increase in individuals migrating from urban to urban areas between provinces) whilst origin urban population has the largest standardised coefficient for urban-rural stream (0.437; a one SD increase in origin urban population is associated with a 0.437 SD increase in individuals migrating from urban to rural areas between provinces), indicating urban population exerts the most substantial push to urban-rural stream in the origin and the greatest attraction for urban-urban migration in the destination with both streams sharing urban origins (Table 4.4).

Among the standardised coefficients of income predictors, those of origin urban and destination rural incomes always remain negative across all four streams, whilst those of origin rural and destination urban incomes are positive for most streams (Table 4.4). Specifically, urban income exerts the most remarkable diverting force to rural-urban stream in the origin (-0.394; a one SD increase in origin urban income is associated with a 0.394 SD decrease in individuals migrating from rural to urban areas between provinces) and the strongest attraction for urban-urban migration in the destination (0.609; a one SD increase in destination urban income is associated with a 0.609 SD increase in individuals migrating from urban to urban areas between provinces), evidencing that the effects of urban income operate in opposite directions in the origins and destinations. Importantly, attractions from destination urban income always outweigh the diverting effects of origin urban income (Table 4.4), which signifies the overall positive association of urban income with all the streams. In terms of rural income, it carries the biggest push force towards urban-urban migration in the origin (0.281; a one SD increase in origin rural income is associated with a 0.281 SD increase in individuals migrating from urban to urban areas between provinces) and the largest diverting effect in the destination (-0.361; a one SD increase in destination rural income is associated with a 0.361 SD decrease in individuals migrating from urban to urban areas between provinces) respectively and simultaneously. Moreover, the diverting effect overshadows the push force for urban-urban migration, indicating the overall negative impact of rural income (Table 4.4). This overall negative impact of rural income also applies to the other three streams (Table 4.4). On the whole, the overall positive impact of urban income contradicts that of rural income (Table 4.4), highlighting the inefficient translation of the rural income growth to employment opportunity increases by comparison.

All standardised coefficients of distance were negative (Table 4.4). Noticeably, the urban-rural stream is most sensitive to changes in distance (-0.432; a one SD increase in distance is associated with a 0.432 SD decrease in individuals migrating from urban to rural areas between provinces), which implies that urban-rural migrants have a general preference for rural destinations close to their urban origins. By contrast, rural-urban migration is the least sensitive to changes in distance (-0.302; a one SD increase in distance is associated with a 0.302 SD decrease in individuals migrating from rural to urban areas between provinces), which is observed in the large scale of urbanisation and industrialisation happening across China (Wan 1995; Zhu 1998). Since rural-urban and urban-rural streams stand as flows and counter-flows between provinces, the size gap of the standardised coefficients of distance between them also illustrates the existence of such asymmetrical patterns within the regional migration system (Slater 1985; Relethford 1986; Schanbacher 2007). In other words, this asymmetrical impact of distance upon flows and counter-flows between provinces re-emphases the importance of developing the original gravity model into its general form to better accommodate the reality, with a statistical investigation into whether this size difference is statistically significant to be conducted below.

Although the comparison of the predictors' standardised coefficients has provided important insights of the predictors' relative effect sizes, it is still not clear whether these effect size differences are statistically significant or not. Under this circumstance, it is necessary to conduct post-estimation testing of the predictors to resolve this problem. By doing so, coefficients of the same predictors tend to be significantly different from each other across all the four equations by using the univariate test where each predictor is tested in turn so its degree of freedom remains as one, evidencing that the four streams are likely to be fundamentally different from each other when separately tested by each predictor in turn. For instance, coefficients of origin urban population are significantly different from each other across all four equations, indicating that they exert distinctive impacts towards out-migration streams of different directions, but origin urban population almost always remains as a significant push factor. Similarly, destination rural population also acts significantly distinctively towards four streams; that is to say, the diverting effects of destination rural population are also significantly different for urban-urban and rural-urban streams, whilst these diverting effects of destination rural population are also significantly different from that of the attraction for urban-rural stream. Needless to say, the attraction for the rural-rural stream is insignificant. As with coefficients of destination urban population, they are significantly different for urban-urban and urban-rural streams, proving that urban-urban and urban-rural streams receive significantly distinctive attractions from destination urban population. Regarding coefficients of origin urban incomes, they are significantly different from each other in urban-urban and rural-urban migration, so it is the case that urban-urban and rural-urban migration streams receive significantly distinctive diverting effects from origin urban incomes. In terms of coefficients of distance, they are significantly different for urban-urban and rural-rural streams, implying that urban-urban and rural-rural streams have significantly different sensitiveness towards migratory distance.

Urban-urban migration stands out with the largest number of predictors with similar coefficients to other stream equations, meaning that urban-urban stream represents the most common characteristics of other streams. On the one hand, for instance, coefficients of origin rural population are not significantly different with each other for urban-urban and urban-rural migration, and neither are those of destination rural income. This implies that urban-urban and urban-rural streams receive similar push forces from origin rural population and similar impediments from destination rural income. On the other hand, urban-urban migration also has a coefficient of destination urban income similar to that of rural-rural streams, whilst its coefficient of origin rural income is similar to those of urban-urban and rural-urban migrations. This shows that urban-urban and rural-rural streams receive similar attractions from destination urban income, and origin rural income exerts similar push forces towards urban-urban and rural-urban migration.

It is of importance to further investigate whether all four equations are significantly different from each other as a whole, though separate tests of predictors have confirmed that four streams are likely to be fundamentally different from each other. There are two reasons accounting for this: for one thing, this multivariate regression model as a whole is built upon the utilisation of the same set of predictors; for another, there is strong correlation among the four responses (correlation coefficients 0.80~0.92, p<0.001) as well as their residuals (correlation coefficients 0.69~0.83, p<0.001). In other words, it is likely that four equations are similar to each other with the same set of predictors and closely related responses, even when separate predictor coefficients are significantly different. A case in point is that urban-urban migration equation seems to have a few predictors with similar coefficients to other three equations. In particular, the one-degree-of-freedom univariate test is designed to test just the simple hypotheses, wherein each time only a predictor can be tested by holding all the other predictors constant across equations. Therefore, there is a real risk in attempting to use univariate tests to test high-dimensional hypotheses, as in this case where it is necessary to conduct the test of all the predictors simultaneously. To resolve this problem, multiple testing procedures are developed to bind the probability of falsely rejecting one of the null hypotheses, wherein p-values of univariate hypotheses are modified to accommodate needs of the high-dimensional hypothesis accordingly (Holm 1979; Simes 1986). By conducting this high-dimensional hypothesis test, it is possible to test all the four stream equations simultaneously as a whole. Importantly, the test result confirms that the four stream equations are significantly different from each other as a whole (Bonferroni-adjusted p-values < 0.001).

However, the residuals of these four stream models are correlated and clustered, though they are normally distributed (Shapiro-Wilk W test: $w = 0.978$, $P < 0.001$). Specifically, residuals of urban-urban and rural-urban migration are most correlated with a coefficient of 0.83, whilst residuals of urban-urban and rural-rural migration are least correlated with a coefficient of 0.69. Furthermore, it is also found that the residuals cluster against both origin and destination provinces simultaneously[78].

4.6 Discussion

This section consists of two sub-sections. Discussion of results for 2000 and 2010 total migration are presented in the first sub-section, whilst the second sub-section details results discussion of 2010's four interprovincial migration streams.

4.6.1 2000 and 2010 total migration

Between 2000 and 2010, the regional migration system went through some dramatic changes, with a growth of the total interprovincial migration from 42 million to 86 million. Based on the modelling

[78] See Figure 4.3 in Chapter 4 of Appendix.

results, a few predicators also behaved remarkably differently in both years. Specifically, destination rural population and income were insignificantly positively associated with the total migration in 2000, whilst these associations became significant and negative in 2010. In other words, it indicates not only that the total in-migration flow became increasingly responsive to destination rural population and income, but also that rural population and income started to act as deterrents in the destination during the period. There is one possible reason for this: the rise of rural income facilitated the rural population mobility (Ye, Wang et al. 2013); and particularly the within-province mobility of the destination rural population increased to such an extent that it stood as a strong competitor to migrants from outside provinces in 2010 (Wu and Yao 2003; Yan 2007). This can be partly proved by the 33%[57] growth in rural income and the growth of the urbanisation rate from 36.2% in 2000 to 49.7% in 2010, though there is no direct within-province migration data for rural population in 2000.

Noticeably, some consistency in the system could also be observed from 2000 to 2010. Most importantly, urban income behaved similarly in both years. Origin urban income was significantly negatively associated with total interprovincial migration in 2000 and 2010, whilst destination urban income remained significantly positively associated with total interprovincial migration in both years by contrast. Moreover, the difference between effect sizes of origin and destination urban income remained large within each model. All of these imply that total interprovincial migration flow remained consistently responsive to (both origin and destination) urban income, and also that urban income acted as a strong diverting factor in the origin and an even stronger pull factor in the destination for total interprovincial migration in both years. Importantly, urban income underwent a radical growth of 21%[57] from 2000 to 2010, which contrasts with its relative stable roles in both years. Then again, part of this contrast can be attributed to the predominate trend of both rural-urban and urban-urban movement in total interprovincial migration, recalling the growth of the urbanisation rate from 36.2% in 2000 to 49.7% in 2010 and the fact that these two streams together constituted 84.0% of the total migration in 2010. The role of urban income is also evidenced in former studies. Specifically, it has been found that destination urban income is a remarkably strong pull factor for both rural-urban and urban-urban interprovincial movements (Wang 2004; Wang and Piesse 2010; Cheng, Nielsen et al. 2014), and that origin urban income greatly facilitates rural-urban and urban-urban migration and is a strong push factor for both streams (Yang 1994; Du and Cheng 2008; Vignoli 2008; Hahn 2010).

4.6.2 2010's four interprovincial migration streams

An examination of migration streams in 2010 found not only that the four streams tend to have significantly different estimates for the same predictors, but also that the four equations are strongly distinguishable from each other as a whole. This implies that these four streams are different from each other in nature, echoing findings from other studies around the world (Vignoli 2008; Kloos, Correa-Oliveira et al. 2010; Kaida and Miah 2015). Indeed, migrants of different streams have different origins and/or destinations, which is caused by distinctive driving forces as evidenced by this chapter. Then again, this can reflect distinctive features and attributes of the four migrant groups to an extent. For instance, urban-urban and rural-urban migrations both arrive at urban destinations, but urban-urban and rural-urban migrants are remarkably different migrant groups and they have significantly different demographic features in terms of age, gender ratio, education, employment and income (Chan and Zhang 1999; Shi, Zheng et al. 2014; Smith-Greenaway and Thomas 2014; Liu 2015). As a consequence, urban-urban and rural-urban streams are remarkable different in the general characteristics of migrants that they contain; for example, rural-urban migration tends to be more seasonal and circular than urban-urban migration (Wang and Fan 2006; Gui, Berry et al. 2012; Rademacher-Schulz, Schraven et al. 2014).

However, three groups of predictors behaved consistently across all four streams in 2010 regarding the direction of their impacts. For the first group, effect sizes of distance for urban-rural and rural-urban streams were similar based on the test results, whilst the hindrance effect of distance remained significant for all streams and total migration. This may be related to the fact that both urban-rural and rural-urban migrants had to overcome some institutional (such as the Hukou system) or physical boundaries between rural and urban areas. For instance, it is generally true that the greater the distance, the more difficult it becomes to make the between-province residential change from rural to urban areas or vice versa not only because of the travelling cost but also because of the Hukou system (Young 2013). One important factor lies in China's regional development policy, which since the 1990s has encouraged localised and within-province urbanisation of small and medium cities (Fan 1995; Fan, Kanbur et al. 2009; Liu, Stillwell et al. 2014). In that sense, distance not only represents the geographical friction, but also acts as a proxy for institutional impacts (Zhang and Tao 2012; Zhang, Zhu et al. 2014). The destination urban population and income were in the second group playing consistent roles across all four equations, and they were always significantly positively associated with all the streams. This finding held true for total interprovincial migration in both 2000 and 2010, highlighting the significant role of the destination urban populations and incomes in the regional migration system in both 2000 and 2010. By contrast, destination rural income was in the third group and remained significantly negatively associated with all four streams in 2010, re-emphasising that its deterrent effects for total interprovincial migration in 2010 also held true across all streams.

Nevertheless, residuals of migration streams in 2010 were strongly correlated and evidently clustered against origin and destination provinces simultaneously, implying dependencies of residuals. In other words, there is still something missing from the above analysis and modelling. One major potential area to improve the prediction precision would be to consider clustering effects in the origin and destination provinces, and this endeavour will be conducted in the next chapter.

4.7 Conclusion

This chapter has offered a new way of using the gravity model to separate out between-province migration flows into their rural and urban components. It has considered the effects of population size, income and distance on urban-urban, urban-rural, rural-urban and rural-rural migration flows in China. It has argued that these four types of flow may differ from each other in terms of their causes and evidence for this has been found in the results that have shown, for example, that the four streams tend to have significantly different estimates for the same predictors and that the four equations are strongly distinguishable from each other as a whole. Nevertheless, the urban-urban migration stream has the biggest potential to represent all the other three migration flows, as equation of urban-urban migration stream has the largest number of predictors with similar coefficients to other stream equations according to post-estimation tests.

Internal population movement trends highlighted here are not unique to China. However, China is special in the sense that its fast urbanisation and industrialisation process has drawn the attention of the academic world to study rural migrants and rural-urban migration, with the underlying assumption that urban areas are more attractive than rural areas mostly because of the income divide (Zhu 2002; Zhou and Zhang 2008; Zhang 2013b). This chapter has challenged this perspective and brought the one-directional interpretation of the rural and urban income divide into question. Indeed, rural areas are not always unfavourable migratory destinations, as 2010 Census has shown that about 1 million undertook the urban-rural movement and that about 12 million made the rural-rural migration. Nevertheless, results from 2010's urban-rural streams also reveal that urban-rural migrants were not entirely driven by economic factors.

This chapter has also highlighted two potential future research directions: one is the use of a wider range of predictors, as the set of variables used here are relatively limited compared with other studies (Fan 2005b; Liu, Stillwell et al. 2014; Shen 2016b); another one is the examination of the clustering effects exhibited in the residuals. The second direction provides the focus for the next chapter, which chooses urban-urban migration stream as the research subject due to its uniqueness and overall representativeness shown by post-estimation tests.

Chapter 5 Analysing interprovincial urban migration flows in China: A new multilevel gravity model approach

The previous chapter has compared total migration between 2000 and 2010 and separated out between-province migration flows in 2010 into their rural and urban components. The results have revealed the clustering effect in the residuals and confirmed that urban-urban stream represents the most common characteristics of other three streams. This chapter is designed to develop a new measurement to account for the clustering effect of the data, which is accomplished by choosing urban-urban migration stream to exemplify the model development process due to its uniqueness and overall representativeness. The uniqueness of the urban-urban migration stream lies in it comprising a large share of total migration (about one third). Given this, it has not received a proportionate amount of research attention. The overall representativeness of urban-urban migration is characterised by it holding the largest number of predictors with similar coefficients to other stream equations based on post-estimation testing in Chapter 4[79].

Please note that we restrict the set of explanatory variables used in this chapter to the key elements of the gravity model (population, income and distance). We do this in order to develop a multilevel extension of the gravity model in an as simple and accessible form as possible. We recommend that researchers who go on to use our proposed model in their own applied work ultimately explore a wider range of covariates.

A paper version of this chapter co-authored with my supervisors Winnie Wang, Richard Harris and George Leckie has been published in *Migration Studies* (Zhang, Wang et al. 2018).

[79] See Sub-section 4.5.2 of Chapter 4 for more details.

5.1 Introduction

This chapter proposes a new method to study between province urban migration flows in China. Migration is a topic of enduring interest in population studies (De Haas 2010a; Molho 2013). The literature shows that migration is a complex social-economic phenomenon exhibiting different features in different geographical contexts. Developed economies recently have seen a rise in counter-urbanisation population movement (Remoundou, Gkartzios et al. 2016), while migration from rural to urban areas has been the predominant trend in the developing world (De Haas 2010b). The focus on rural-to-urban and urban-to-rural migration (Ezra and Kiros 2001; Fan and Wang 2008; Remoundou, Gkartzios et al. 2016) relates to the classic two-sector migration theory whereby labour transfers from the primary and rural employment sectors to the secondary and urban sectors at the beginning stage of industrialisation (Lewis 1954b; Harris and Todaro 1970; Todaro Michael 1976), and also to the post-industrialisation migration framework, which emphases urban-to-rural movement and population decentralisation processes away from cities into less densely populated areas (Berry 1976; Vartiainen 1989).

Migration between urban areas (i.e. urban-urban migration) is less studied perhaps because it fits into neither framework. It is, however, prevalent in both developed and developing worlds. In industrialised countries where urbanisation is reaching its saturation level, the majority of movements are between urban areas. Urban-urban migration has been studied as part of the dynamics of urban systems (Andersson, Haag et al. 2012) or as a component contributing to differential urban growth (Pumain and Sanders 2013). In addition, urban-urban migration has been studied as a spatial movement between different labour markets most closely linked with economic factors such as employment and income (Flowerdew and Salt 1979; Poot 1986) but also with amenities (Greenwood and Hunt 1989) and house prices (Johnston, Owen et al. 2016).

In the developing world, while rural-urban migration is the dominant trend, urban-urban migration has become the main form of population movement in Latin America since the 1980s due to its accelerated urbanisation process (Cerrutti and Bertoncello 2003). Generally, however, there have been limited attempts to examine how the pattern and process of urban-urban migration vary from rural-urban migration (Machado and Hakkert 1988; Shefer and Steinvortz 1993). Indeed, little is known about what may be the world's largest urban-urban migration, which is occurring in China and is caused by rapid urbanisation with lessening institutional restrictions and rising population mobility (Feng 2002; Wu and Yao 2003). There has been some discussion of how urban-urban migrants integrate with mainstream society and access the welfare system in a few cities (Yang 2013; Cheng, Nielsen et al. 2014), and on the links between increasing urbanisation, growing urban-urban migration and the career-driven characteristics of urban-urban migrants (Vignoli 2008; Hahn 2010). Nevertheless, broader understanding of the macro-level urban-urban migration patterns and its mechanisms are unclear. This is despite China's census reporting 260 million people to have migrated internally in 2010 (approximately 20% of the total population), a third of which were urban-urban migrants.

This chapter addresses two gaps in the literature: the lack of attention given to urban-urban migration in China, and the lack of an appropriate statistical model to do so. The study proposes a multilevel gravity model of migration which combines the merits of the linear regression formulation with multilevel modelling to investigate: (a) how flows originating from the same Chinese province and (b) how flows ending at the same province vary from each other in regard to the average number of migrants they contain; (c) what is the correlation between the average out-migration and in-migration flows across provinces; and (d) how the reciprocal flows between two provinces are related. We illustrate the model using data about urban-urban migration flows within China obtained from the 2010 Chinese census.

5.2 The gravity model, multilevel modelling and inter-flow dependencies

The gravity model is widely used in analysing migration flows (Fan 2005b), where the numbers of people moving between locations are modelled as a function of the attributes of the locations such as population size and GDP, and of the physical or socio-economic distance between places (Converse 1949; Christian and Braden 1966).

There are several reasons for its popularity; amongst them are that the gravity model is capable of incorporating both origin and destination attributes when modelling migration flows (Beine, Bertoli et al. 2014). The model also is flexible in allowing predictors of migration to be added beyond the original form of the model where only distance and populations are used (Beine, Bertoli et al. 2014; Shen 2015). In China's context, among the existing gravity model studies, some have been conducted to determine how the total migration flows distribute across space and over time (Fan 2005b), and others focus on the determinants of the total migration flows (Shen 2012).

Conventionally gravity models are formulated and estimated as linear regression models, where each row of the data matrix represents a tally of movements from one place to another – a flow – and these flows are tacitly assumed to be independent of one another. However, Origins and destinations can be connected in four key ways as illustrated in Figure 5.1: Type 1, indirectly when multiple destinations receive migrants from the same origin (e.g., flows 1 and 6); Type 2, when multiple origins send migrants to the same destination (e.g., flows 2 and 5); Type 3, when an origin sends to a destination, which is it itself a destination to another origin (e.g., flows 3 and 1); and Type 4, directly when there is migration in both directions between two places, so both places are simultaneously an origin and a destination to the other (e.g., flows 1 and 2). These connections suggest that the flows between places are not independent of one another. However, these four flow dependencies are seldom addressed in the regional migration literature. If the assumption of independence is invalid and there are dependencies between the flows then the estimates of statistical significance and of effect size are affected (the former typically are over-stated, whereas the latter may have deviated from their true value). In fact, some degree of dependency is almost inevitable: for a gravity model, each row in the data matrix provides an origin, a destination and the number of people that moved between them, as well as other attributes of the places that may explain the flow. Unless those origins and destinations are all unique, then those attributes are necessarily repeated, creating group dependencies which ought to be controlled for.

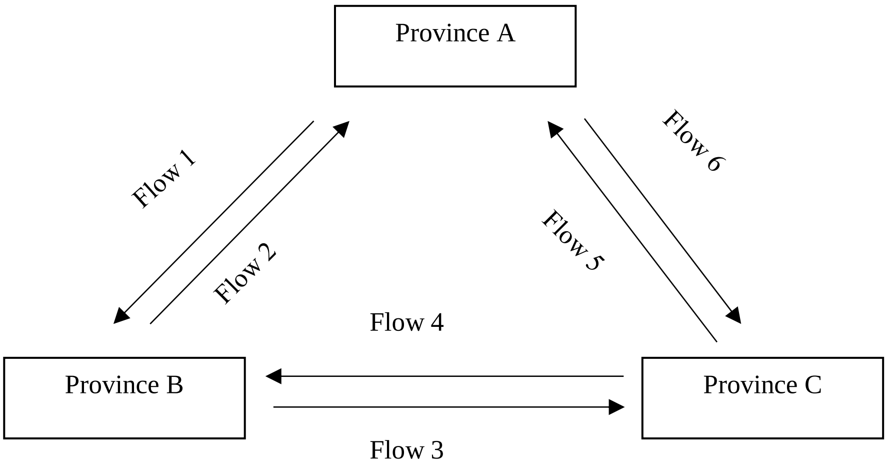

Figure 5.1 Illustration of the different potential population flows between three provinces and how they are interrelated

Aside from the potential for estimation errors, standard linear regression lacks the capacity to quantify the strength of the inter-dependencies between the flows. Multilevel modelling can be seen as a generalisation of linear regression, designed to deal with hierarchically structured or cross-classified data (Goldstein 2011; Leckie 2013; Harris 2017), which has the potential to estimate similarities between observations that belong to a common group – a common origin, for example. Only a small number of migration studies have used multilevel modelling techniques. Amongst those that have, some examined the determinants of migration among certain migrant groups at different geographical scales (Kallan 1993; Ezra and Kiros 2001), whilst others investigated interregional flows but treating in-migration and out-migration as independent events (Dennett and Wilson 2013). In the context of China, Yang and Guo (1999) use multilevel modelling to examine gender differences in the determinants of labour migration at the individual/household and community levels. While Shen (2016a) uses a two-stage Poisson version of the gravity model with a network spatial filter and decomposes the estimation errors into the overall effect of the constant, of the relative emissivity of the origin and the relative attractiveness of the destination, and a measure of interaction between places.

What has not systematically been examined is the potential connections between origins and destinations – the ways that the flows are related to and dependent upon one another. A partial exception was undertaken by (Thomas, Stillwell et al. 2015), who used multilevel modelling to allow for individual and contextual variations by origin and by destination in the distances moved by residential migrants in England and Wales, but did not consider the correlations between origins and destinations in terms of the migration flow.

In this chapter, we are interested in allowing for and quantifying the four types of flow dependency which arise when two flows share provinces in common. First, two flows may correlate due to sharing a common origin (Type 1). Second, two flows may correlate due to sharing a common destination (Type 2). Third, two flows may correlate if the destination in the first flow is the origin in the second flow (Type 3). Fourth, two flows may correlate if there are reciprocal flows between two places (Type 4). These dependencies are in part induced by unmodeled origin and destination effects. Origin effects reflect variation between places in the number of migrants that move away. Destination effects reflect variation between places in terms of the volume of migrants that they attract. These two sets of effects may well be associated, reflecting a correlation between the in-migration and out-migration flows for a place when it both sends out and receives migrants. Even if we take into account these correlated origin and destination effects the residuals may well continue to be correlated within pairings of provinces, reflecting the bilateral flows between themselves. All these four effects represent the spatial dependencies between places – how the numbers of migrants sent or received at one place can be dependent on the number of migrants sent or received at others. Instead of treating space as a vacuum in which all that matters is the mass of attraction between any two places that are somehow isolated from other places, we instead adopt a system-wide perspective and allow for the inter-connections between places through migration flows. Although the importance of spatial dependency has been stressed in multiple fields of social science (Fingleton 1986; Getis 1990; Leorato and Mezzetti 2016), it has not received comparable attention in migration studies. To fill this empirical gap, we propose multilevel modelling extensions to the traditional linear regression formulation of the gravity model of migration.

5.3 Methodology

We start this section by presenting three increasingly realistic implementations of this model: Model 1, the traditional linear regression formulation of the gravity model[80] (Equation 3.3); Model 2, a standard cross-classified multilevel model formulation which extends Model 1 to capture systematic variation in out- and in-migration across provinces; and Model 3, an extended version of Model 2 where we additionally allow for correlations in the out and in-migration flows. We then show how only Model 3 captures the four dependencies that arise in migration data. Finally, we discuss estimation and the illustrative data.

In addition, 5 more intermediate models[81] are developed to enhance the model development process and test the robust of Model 3. Model 1a and 1b respectively investigate the origin- and destination-clustering effect on the basis of Model 1 (the traditional linear regression formulation of the gravity model). Based on the standard cross-classified multilevel model formulation of Model 2, Model 2a controls the origin- and destination-variance to be equal, whilst Model 2b allows for the correlation of

[80] See Sub-section 3.2.3 for more details.
[81] Their results are presented in Section 4.7 of Appendix.

origin and destination effects. The final intermediate model, Model 3a, is established by further holding the origin- and destination-variance to be equal in Model 3.

5.3.1 Model 2: The standard cross-classified multilevel formulation

A fundamental limitation of the linear regression model is that it assumes the residual migration flows ε_{ij} are independent. However, we expect residual migration flows to systematically vary across origins and destinations. Specifically, we expect the out-migrations from a given province to be positively correlated as they share a common origin. Likewise, we expect the in-migrations to a given province to be positively correlated as they share a common destination. Linear regression ignores these dependencies and will therefore estimate spuriously precise regression coefficients raising the risk of type I errors of inference: we might conclude covariates to be significant when they are not. The estimated parameters may also be ‘unstable’ and deviate from their true values.

We propose a multilevel modelling based approach to dealing with the complex residual dependencies which arise when modelling migration flows. However, before we introduce this approach, we note that there are other general approaches to dealing with simpler forms of clustered data which could potentially also be applied to migration data. In particular, one might attempt to account for the Type 1 and 2 dependencies introduced above by replacing the model-based standard errors in the usual linear regression formulation of the gravity model with their two-way cluster-robust counterparts (Cameron, Gelbach et al. 2011). However, this approach does not straightforwardly allow for the more nuanced Type 3 and Type 4 dependencies we aim to capture. Furthermore, the two-way cluster-robust standard error approach does not additionally quantify and therefore allow one to substantively interpret the magnitudes of these four forms of dependencies, nor does it allow one to make predictions regarding specific province origin and destination effects. A central argument in this chapter is that both these lines of investigation are substantively insightful when studying migration flows.

We address this concern by specifying a multilevel version of the model which includes cross-classified origin and destination random effects to account for systematic residual variation in out-migration and in-migration across provinces. Model 2[82] can be written as

$$y_{ij} = \beta_0 + \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + \mathbf{x}'_{2j}\boldsymbol{\beta}_2 + \mathbf{x}'_{3ij}\boldsymbol{\beta}_3 + o_i + d_j + e_{ij}$$

$$o_i \sim N(0, \sigma_o^2)$$

$$d_j \sim N\left(0, \sigma_d^2\right)$$

$$e_{ij} \sim N(0, \sigma_e^2) \qquad (5.1),$$

where o_i and d_j denote the origin and destination random effects and e_{ij} the revised residual.

The random effects and residuals are typically stated to be normally distributed with zero means and constant variances. However, we note that normality of the random effects and residuals are not required for consistent estimation of the model parameters and standard errors. It should be kept in mind that empirical Bayes predictions of the random effects do rely on at least approximate normality and so we recommend one checks this assumption when we apply these models to the data. The origin and destination variances σ_o^2 and σ_d^2 quantify the degree to which origins and destinations vary in average out-migration and average in-migration having adjusted for the covariates. The residual variance σ_e^2 quantifies the remaining variation.

Dividing each variance component by the total residual variance $\sigma_o^2 + \sigma_d^2 + \sigma_e^2$ gives variance partition coefficients (VPCs) which can be used to quantify the relative importance of origins and destinations

[82] This model is conceptually similar to the origin/destination constrained gravity model estimated by fixed effects (Guy 1987). What is different here is that the origin and destination effects are specified as random effects. This version of the model has the considerable benefit of being able to include origin and destination covariates. Further, this version of the model also allows for evaluating how unexplained origin and destination differences change as origin and destination covariates are included.

in explaining residual migration. For instance, Thomas, Stillwell et al. (2015) used VPCs to estimate the distances moved by residential migrants in England and Wales and found that city-region random effects are more important than the neighbourhood random effects. The VPCs allow explanation of whether there are unexplained differences between provinces in terms of the numbers of migrants they send or receive or whether the differences between what occurs and what the model predicts is simply random between the individual flows with no evidence of place effects.

5.3.2 Model 3: The extended cross-classified multilevel formulation

Model 2 is an improvement on the standard regression but still assumes that provinces' origin and destination effects are independent of one another. However, a province's level of out-migration is likely to be linked to their level of in-migration, even after adjusting for the covariates. For example, provinces which in general exhibit higher than expected out-migration might be expected to exhibit lower than expected in-migration and vice versa. Put simply, we might expect variation in net-migration over and above that predicted the characteristics of the provinces captured by the covariates. In Model 3, we therefore allow the origin and destination random effects to correlate, $\text{Corr}(o_i, d_i) = \rho_{od}$.

Model 2 also assumes that residual migration flows are independent within each pair of provinces. However, here too we might expect a systematic relationship. Namely, where there is a higher than expected flow from one specific province to another specific province we may see a higher than expected flow in the return direction. That is, it seems likely that we might see particular province parings which exhibit higher (or lower) than expected migration flows in both directions. In Model 3, we therefore also allow for correlated within province-pair residuals, $\text{Corr}\left(e_{ij}, e_{ji}\right) = \rho_{ee}$.

Model 3 can therefore be written as

$$y_{ij} = \beta_0 + \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + \mathbf{x}'_{2j}\boldsymbol{\beta}_2 + \mathbf{x}'_{3ij}\boldsymbol{\beta}_3 + o_i + d_j + e_{ij}$$

$$\begin{pmatrix} o_i \\ d_i \end{pmatrix} \sim N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_o^2 & \\ \sigma_{od} & \sigma_d^2 \end{pmatrix}\right\}$$

$$\begin{pmatrix} e_{ij} \\ e_{ji} \end{pmatrix} \sim N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & \\ \sigma_{ee} & \sigma_e^2 \end{pmatrix}\right\} \qquad (5.2),$$

where the origin-destination and residual correlations can be derived from the associated variance and covariance parameters in the usual way, $\rho_{od} = \sigma_{od}/\sigma_o\sigma_d$ and $\rho_{ee} = \sigma_{ee}/\sigma_e^2$.

We further specify Equation (5.2) as $\mathbf{x}_i' = [\ln(p_i)\ \ln(I_i)]$ and $\mathbf{x}_j' = [\ln(p_j)\ \ln(I_j)]$, where I_i and I_j are the incomes of the origins and destinations, in line with what is commonly used in most studies and reflecting the importance of economic factors in the migration system (Fan 2005b; Beine, Bertoli et al. 2014). The origin population p_i is used to represent the migration potential of the sending place. According to the International Organization for Migration (Laczko, Tjaden et al. 2017), migration potential is a valuable indicator of the actual migration flows that take place. The total number of individuals who are ready to migrate is proportional to the size of the population in the sending area. So the larger the origin population, the greater migration potential a place will have. In contrast, the destination population p_j provides a strong proxy for the employment prospects a place has. According to neoclassical economics (Massey, Arango et al. 1993), migrants tend to move to places with more favourable employment conditions. A larger population suggests a bigger and more diverse labour market which provides more and varied employment opportunities. Therefore, it will be much easier for migrants with different specialised skills to find work in destinations with larger populations. This is especially true for speculative migrants (Gordon 1995).

We note that Model 3 takes the same form as the social relations model (Kenny and Kashy 2011), which has recently been adopted to handle counts response variables (Koster and Leckie 2014) and will be of interest to researchers who prefer working with Poisson frameworks (Shen 2016a).

Table 5.1 Equations for the model-implied correlations conditional on the covariates between flows corresponding to the four key dependencies shown in Figure 5.1

Dependency	Correlation	Model 1	Model 2	Model 3
Type 1	$\mathrm{Corr}(y_{ij}, y_{ik})$	0	$\frac{\sigma_o^2}{\sigma_o^2+\sigma_d^2+\sigma_e^2}$	$\frac{\sigma_o^2}{\sigma_o^2+\sigma_d^2+\sigma_e^2}$
Type 2	$\mathrm{Corr}(y_{ij}, y_{kj})$	0	$\frac{\sigma_d^2}{\sigma_o^2+\sigma_d^2+\sigma_e^2}$	$\frac{\sigma_d^2}{\sigma_o^2+\sigma_d^2+\sigma_e^2}$
Type 3	$\mathrm{Corr}(y_{ij}, y_{jk})$	0	0	$\frac{\sigma_{od}}{\sigma_o^2+\sigma_d^2+\sigma_e^2}$
Type 4	$\mathrm{Corr}(y_{ij}, y_{ji})$	0	0	$\frac{2\sigma_{od}+\sigma_{ee}}{\sigma_o^2+\sigma_d^2+\sigma_e^2}$

Note. See Sub-section 5.6 of Appendix for derivations.

Table 5.1 presents equations for the model-implied correlations between migration flows corresponding to the four key dependencies that we have identified in migration data[83]. Type 1: two flows may correlate due to sharing a common origin; Type 2: two flows may correlate due to sharing a common destination; Type 3: two flows may correlate if the destination in the first flow is the origin in the second flow; Type 4: two flows may correlate if they are reciprocal flows between the same two places. Model 1 (Equation 3.3), the linear regression formulation implicitly assumes all these correlations to be zero, and therefore ignores all four dependencies. Model 2 (Equation 5.1), the standard cross-classified multilevel formulation allows us to estimate the first two correlations and therefore allows for the first two types of dependency. Only Model 3 (Equation 5.2), our extended cross-classified multilevel formulation with correlated origin and destination effects and correlated within province pair residuals, allows us to estimate all four correlations and therefore allow for all four forms of dependency.

[83] See Sub-section 5.6 of Appendix for derivations.

We fit the models by iterative generalised least squares (equivalent to maximum likelihood estimation) using MLwiN 2.36 where we call MLwiN (Rasbash, Charlton et al. 2009) from within Stata 14 using the user-written runmlwin command (Leckie and Charlton 2013).

5.3.3 Data

Data used in this study are mainly drawn from China's 2010 Census (migration and population data) and 2011's China Statistical Yearbook (income data). A summary of the data used is given in Table 5.2. Each variable is log-transformed in the models but is shown in its original scale in Table 5.2. Origin-destination distance is calculated as the distance between provincial capital cities, whilst urban income is defined as per capita disposable income of urban households[84].

Table 5.2 The response variable and covariates used for the analysis

Original data	Level	Units	Observations	Mean	SD	Min	Max
Urban-urban migration	Flow	00s persons	930	1447.80	3551.79	1	48921
Urban population	Province	Millions of persons	31	12.40	7.06	0.44	29.58
Urban income	Province	000s yuan	31	18.07	4.70	13.19	31.84
Distance	Flow pair	km	465	1379.53	729.31	113.69	3598.79

5.4 Results

Table 5.3 shows the results from models 1, 2 and 3: the linear regression, standard cross-classified multilevel and extended cross-classified formulations of the gravity model, respectively. All three models include the natural log of the origin and the destination populations and incomes as covariates as well as the natural log of the distance between each pair of provinces. Recall that Model 2 extends Model 1 by introducing the origin and destination random effects, while Model 3 further allows for the

[84] The detailed data acquisition process is in Section 3.1 of Chapter 3.

effects to correlate. Likelihood ratio tests show that Model 3 is significantly preferred to Model 2 ($\chi_2^2 = 406.1, p < 0.001$) which in turn is preferred to Model 1 ($\chi_2^2 = 316.4, p < 0.001$).

Additional likelihood ratio tests of intermediate models confirm that Model 3 is also significantly preferred to Model 1a ($\chi_3^2 = 643.6, p < 0.001$), Model 1b ($\chi_3^2 = 482.0, p < 0.001$), Model 2a ($\chi_3^2 = 320.1, p < 0.001$), and Model 2b ($\chi_1^2 = 315.6, p < 0.001$), whilst the more complicated model of Model 3a is not significantly preferred to Model 3 ($\chi_1^2 = 3.8, p > 0.05$). Results of models 1, 2 and 3 are presented and discussed in the main body of this chapter[85].

In all three models the estimated coefficients are in the expected directions, similar in magnitude across models, and statistically significant at the 0.1% level (Table 5.3). The models show that the larger the population of the origin and/or the destination, the greater the flow between them. Specifically, a 10% increase in the origin population is associated with an approximate 9% increase in out-migration, all else equal, while a 10% increase in the destination population is associated with an approximate 6% increase in in-migration. Origins with lower income send out more migrants and destinations with higher urban income attract stronger migration streams in accordance with the neo-classical economy migration theory (De Haas 2010a). Specifically, a 10% decrease in the origin income is associated with an approximate 11% increase in out-migration while a 10% increase in destination population is associated with an approximate 21% increase in in-migration. As expected the distance between the provinces acts as an impeding factor (the longer the distance the weaker the flow).

[85] Discussion of intermediate model results presented in Section 5.7 of Appendix.

Table 5.3 Results from linear regression and multilevel formulations of the gravity model of migration

	Model 1		Model 2		Model 3	
Variables	Est.	SE	Est.	SE	Est.	SE
Fixed part						
Constant	4.995*	0.613	7.534*	1.396	7.410*	1.503
Log of origin urban population	0.921*	0.032	0.909*	0.075	0.909*	0.074
Log of destination urban population	0.632*	0.032	0.619*	0.106	0.620*	0.105
Log of origin urban income	-0.990*	0.122	-1.068*	0.284	-1.064*	0.283
Log of destination urban income	2.182*	0.122	2.104*	0.403	2.108*	0.401
Log of distance	-0.818*	0.044	-1.105*	0.039	-1.091*	0.051
Random part						
Origin province variance			0.106*	0.030	0.104*	0.030
Destination province variance			0.224*	0.060	0.221*	0.060
Individual flow variance	0.641*	0.030	0.348*	0.017	0.348*	0.021
Origin-destination correlation					0.105	0.193
Flow-pair correlation					0.719*	0.023
Dependency (correlation conditional on the covariates)						
Type 1: Common origin			0.156*	0.040	0.155*	0.040
Type 2: Common destination			0.331*	0.062	0.328*	0.062
Type 3: Destination in first flow is origin in the second					0.024	0.044
Type 4: Reciprocal flow (shared origin and destination)					0.419*	0.090
Deviance	2225.8		1819.7		1503.3	

Note: Response variable is the log migration flow (in 1000s). Est. = Estimates. SE = standard errors. * denotes p < 0.001.

While the magnitude of the coefficients is similar across models, it is important to note that the standard errors differ dramatically. Moving from Model 1 to Model 2, the standard errors of the province level covariates (incomes and populations) approximately double when we take into account the clustering of migration flows by origins and destinations. The smaller standard errors in Model 1 are therefore spuriously precise illustrating that the standard linear regression formulation of the gravity model is inadequate for modelling migration flows with shared origins and/or destinations. Moving from Model 2 to Model 3 sees no further change to the standard errors of the province-level covariates; rather it is now the standard error of the flow-pair level covariate (distance) which increases (by approximately 30%) when we additionally take into account the correlation between reciprocal flows. Thus, where interest lies in flow-pair level covariates, even Model 2 the standard cross-classified multilevel model proves insufficient.

For Model 3, the origin, destination and residual VPCs account for 16%, 33% and 52% of the total residual variance, respectively. Thus, having adjusted for the covariates, we see that provinces vary far more in the number of migrants they attract than in the number of migrants they send; destination effects vary more than origin effects. Nonetheless, half the variation in migration flows unexplained by the covariates cannot be attributed to origin and destination effects and instead relates to the unique interactions and relationships between pairs of provinces.

The estimated origin-destination correlation of 0.11 is small and not significant and so it is not the case that provinces that exhibit unusually high out-migration also exhibit unusually high or low in-migration. In contrast, the estimated flow-pair correlation of 0.72 is large and significant, suggesting that where one province sends a higher than predicted number of migrants to another, we in general also see a higher than predicted number of migrants sent from another province. Reciprocity in flows between provinces is clearly an important feature of urban-urban migration but this would have gone unnoticed in Model 2.

Table 5.3 presents the estimated correlations conditional on the covariates for the four forms of dependency, which further confirms findings of the random effects. Specifically, the model-implied correlations of flows sharing a common origin (Type 1) is 0.15, whilst that of flows sharing a common destination (Type 2) is 0.33 (more than twice of that of Type 1). However, the correlation between two residual flows where the destination of the first flow is the origin of the second (Type 3) to be just 0.02 and insignificant, and the correlation between reciprocal residual flows (Type 4) is 0.42.

Figure 5.2 plots the residual differences between the origins. They are shown in the original measurement units by exponentiating the predicted origin random effects and their 95% confidence limits for each province, which is then compared to the reference line (Origin province effect=1). Results also show that the predicted province origin (and destination) random effects are approximately normally distributed. Noticeably, the original reference line is ‘Origin province effect=0’, which represents the theoretical mean of the normally distributed residuals. After exponentiating, the reference line of Figure 5.2 becomes ‘Origin province effect=1’, still representing the exponentiated theoretical mean but now being proportional to the overall average number of out-migrants across all provinces. The reason to do this is to make it more explicit to interpret the origin province effects on the original scale, as the data is log-transformed in the model. That is to say, the unit of the origin effects is thousand, as the original data unit of urban migration is hundred and associated with a multiplier of 10 due to the 10% sampling procedure. For instance, the origin random effect of Chongqing has the mean of 1.26 and an interval between 1 and 1.58 (Figure 5.2), which does not overlap with the reference line and means that Chongqing significantly systematically sends 260 more migrants on average than the overall national mean. In a similar way, based on whether the 95% confidence intervals overlap the reference line or not, the provinces have been put into three groups. The first group contains provinces with above average residuals where the confidence intervals do not overlap with the overall average, indicating that they depart significantly from the theoretical gravity model by systematically exporting more urban-urban migrants than predicted by their population, income and distances to other provinces. The five provinces are Zhejiang, Fujian, Ningxia, Heilongjiang and Chongqing represented by the black dots. The second group contains provinces that are significantly below average and systematically export fewer migrants. They are Shanxi, Yunnan, Guangxi and Guizhou represented by the light grey dots. The remaining 22 provinces do not appear to have origin effects that deviate significantly from the overall average.

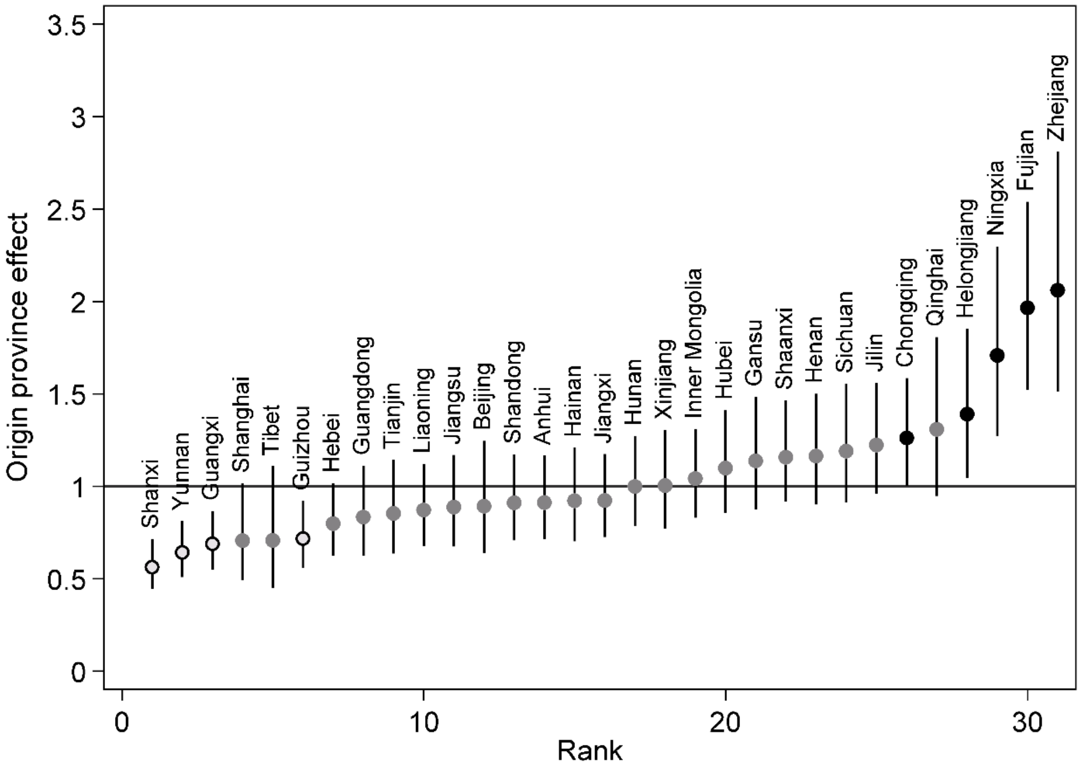

Figure 5.2 Predicted province origin random effects plotted in rank order with 95% confidence intervals (unit: thousands)

Figure 5.3 maps the residuals using the same colour scheme introduced in Figure 5.2, which indicates a positive spatial autocorrelation overall with a marginally significant[86] Moran's I value[87] of 0.15 (Expected value[88] = -0.03, p = 0.09) . That is, Figure 5.3 may show the spatial clustering of provinces with similar origin effects in some way. For instance, three neighbouring provinces (Yunnan, Guizhou

[86] The global Moral I is related to sample size (Huitema, 1991). The insignificance here may be relevant to the small number of sample size (31 provinces).

[87] The calculation of the global Moral I is $I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (o_i - \bar{o_i})(o_j - \bar{o_i})}{\sum_i (o_i - \bar{o_i})^2}$ (Moran, 1950). N is the number of spatial units (provinces), o_i and o_j are the residuals of origin province i and j respectively, and $\bar{o_i}$ is the mean of o_i. The spatial weight matrix w_{ij} is constructed based on the common definition of neighbours, wherein a weight of 1 is given if two provinces i and j are neighbours, and 0 otherwise. By definition, w_{ij} equals 0, when i equals j.

[88] The expected value of Moran's I under the null hypothesis of no spatial autocorrelation is -0.03, calculated by $E(I) = \frac{-1}{N-1}$ (Moran, 1950). N equals 31 here, as there are 31 provinces in the analysis.

and Guangxi) with below-average exporting capabilities cluster at the south-western corner, two neighbouring coastal provinces (Fujian and Zhejiang) of above-average exporting abilities agglomerate in the southeast, whilst the majority of the provinces with average exporting capabilities form the biggest clustering in the map.

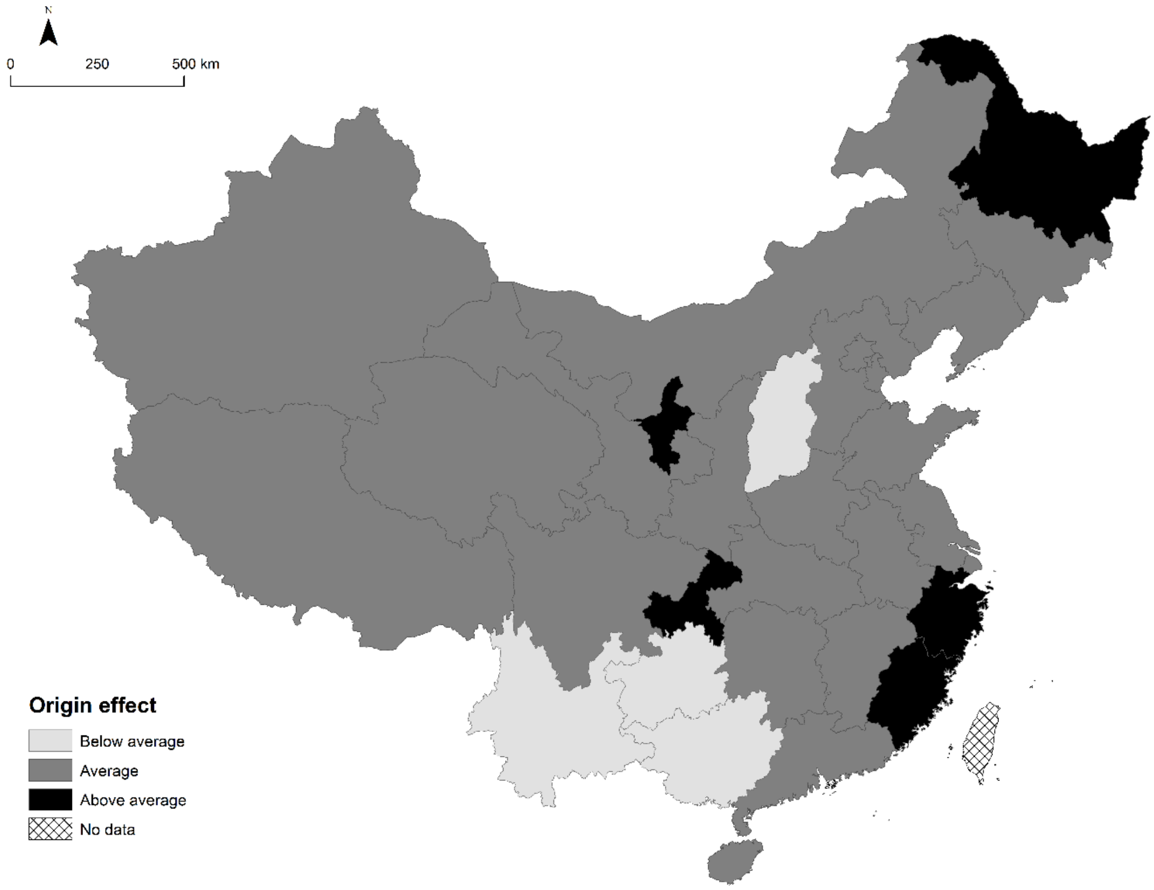

Figure 5.3 Spatial pattern of the predicted province origin random effects

The destination effects can be considered in the same manner as the origin effects. In Figure 5.4, there are nine provinces receiving significantly more migrants above the average (Xinjiang, Hainan, Guangdong, Beijing, Shaanxi, Ningxia, Gansu, Sichuan and Yunnan) and eight provinces that receive significantly less than average (Tibet, Henan, Anhui, Shanxi, Hunan, Tianjin, Jiangxi and Inner

Mongolia). That the number of significant destination effects exceeds the number of significant origin effects is expected as destinations were shown to be twice as variable as the origins.

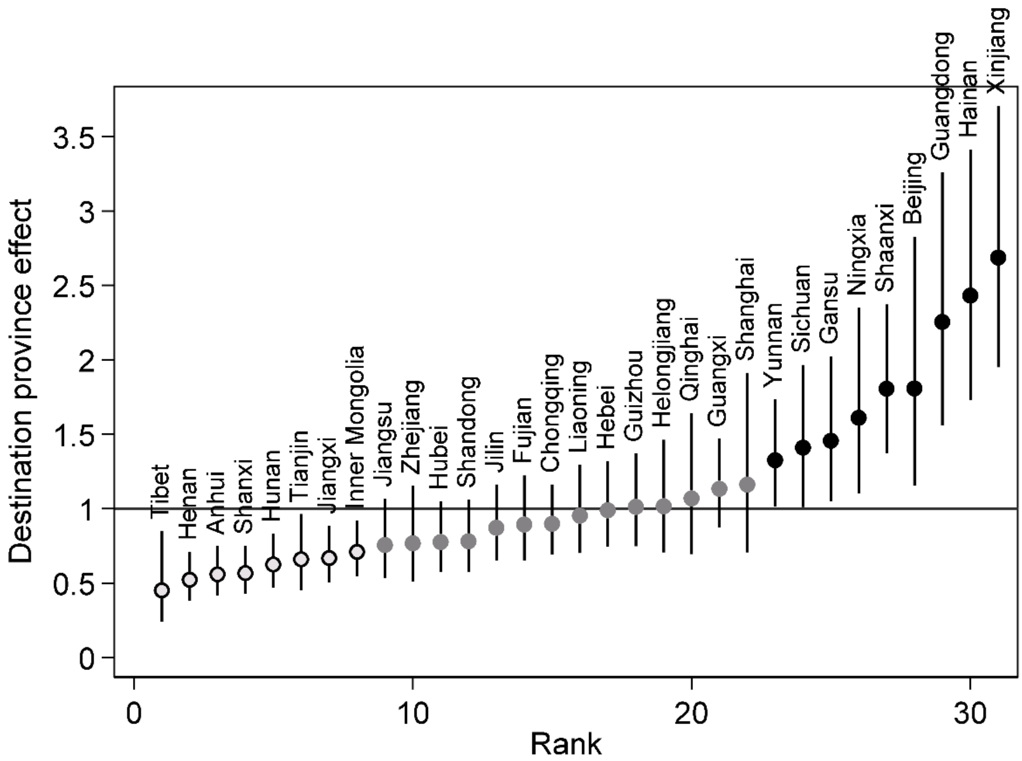

Figure 5.4 Predicted province destination random effects plotted in rank order with 95% confidence intervals (unit: thousands)

Mapping the destination effects, as shown in Figure 5.5, may reveal some spatial variations for the destination effects[89] (Moran's I = 0.03 (Expected value = -0.03), p = 0.53). In general, more spatial

[89] Again, the insignificance here may be relevant to the small number of sample size (31 provinces). The calculation of the global Moral I for the destination effect is conducted in the same way for the origin effect, with the equation $I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (d_i - \overline{d_i})(d_j - \overline{d_j})}{\sum_i (d_i - \overline{d_i})^2}$ (Moran, 1950). d_i and d_j are the residuals of destination province i and j respectively, and $\overline{d_i}$ is the mean of d_i. The spatial weight matrix w_{ij} is constructed in the same way as the calculation of the global Moral I for the origin effect. The expected value remains the same as -0.03. Because the number of the spatial units (31 provinces) does not change in the analysis.

variations and more complicated spatial patterns can be observed in Figure 5.5, with ribbon-like clusters and heterogeneous spots scattering all over the map.

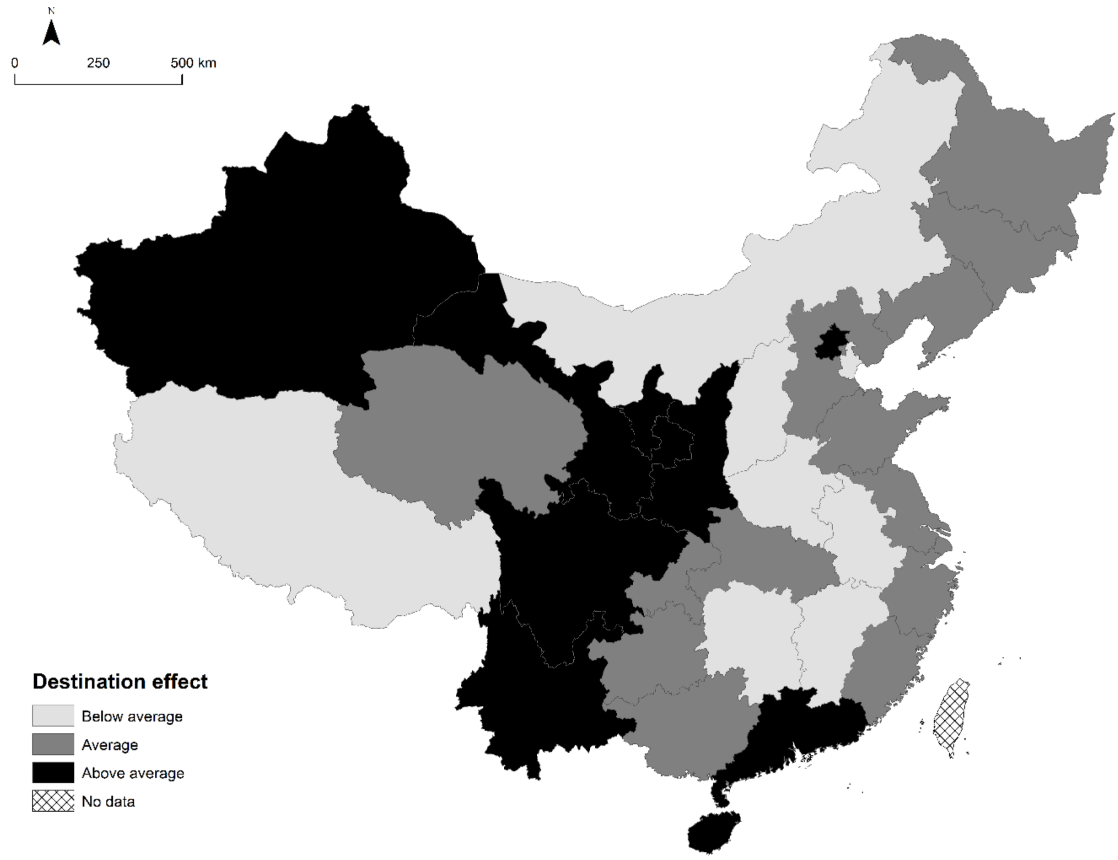

Figure 5.5 Spatial pattern of the predicted province destination random effects

5.5 Discussion

While rural-urban migration has remained important and has prevailed in developing countries, urban to urban migration has started to gain momentum as the developing world becomes more and more urbanised. This trend has already been observed in Latin American countries such as Brazil (Machado and Hakkert 1988), Mexico (Lozano-Ascencio, Roberts et al. 1996), and Colombia (Shefer and Steinvortz 1993). China's urbanisation level has risen rapidly over the past three decades growing from 21% in 1982 to 56% in 2016. In 2010, around 90 million people migrated between urban areas in China.

Identifying factors that affect inter urban migration flows helps shed light on the migratory process that urban-urban streams exhibit.

The results from Model 3 echo previous findings to some extent, namely the effect of origin, destination populations and incomes and distance on total migration (Fan 2005b; Shen 2012; Liu, Qi et al. 2015). For urban populations, those of both origin and destination have significant positive effects upon urban-urban migration but that of the destination is overshadowed by its origin counterpart. Even though previous studies on total migration in China have shown similar results that the origin population is a more influential indicator for migration flows than the destination population (Fan 2005b), this finding is a little counterintuitive for urban-urban migration. This perhaps is due to the fact that our study is using the urban population of the entire province instead of a particular city and therefore a larger urban population may not necessarily translate into the effect of a major urban labour market if the urban population is distributed across many medium and small size cities. Nevertheless, our models show that larger urban populations at both origin and destination provinces help to generate greater migration flows between them which agrees with existing findings relevant to migration stock theory (Fan 2005a, 2005b; Shen 2015).

In terms of urban income, that of destination exerts significant positive impacts upon urban-urban migration, whereas origin urban income has strong negative effects by contrast. This result is consistent with the conventional push-pull framework: migrants are pushed out of areas with lower income and attracted to areas with higher earning. The gap between the effect sizes of origin and destination urban income signposts that the economic pull force plays a much bigger role than push factors in this migration system which agrees with the general observations that pull factors outweigh push factors in most migration flows, particularly for economic migration. This also suggests that the urban-urban migration in China is mainly economically driven.

With respect to distance, it has a substantial negative effect upon urban-urban migration flows: as the distance between the provincial capital cities increases by 10%, the migration stream reduces by 11%. This is surprising given the substantial investment China has made in domestic infrastructure and transportation and the associated reductions in travelling cost which this has brought about in recent decades (Luo, Zhu et al. 2014). Still, the deterrence effect of distance on urban-urban migration flow is consistent with existing findings (Yan 2007; Shen 2013).

What the multilevel model shows, which standard approaches cannot, is that all the random effect parameters, except for the origin-destination correlation coefficient, are significant and contribute substantially to explain the interprovincial urban-urban migration flow residuals. Adding origin and destination variances as well as allowing flow-pairs to correlate greatly improves the fit of the model.

Residual differences in the number of migrants leaving the provinces seem to be closely linked with the provinces' urbanisation level. Except for Ningxia, provinces with significantly higher than average exporting capabilities all had an above the national average urbanisation level (50%) in 2010 ranging from 62% in Zhejiang to 53% in Chongqing. Even Ningxia's urbanisation level was only slightly lower than the average with 48%. Provinces with significantly lower exporting capabilities all had below average urbanisation level. In fact, other than Shanxi (48%), the remaining provinces were among the least urbanised provinces in China. Guizhou, Yunnan, Gansu and Guangxi were ranked near the bottom in terms of urbanisation, and are all located in less developed Western China. This finding lends support to a possible hypothesis raised in the previous section that even though origin urban population helps increase migration flow, more urbanised provinces may provide more opportunities and therefore do not have such a strong push force as less urbanised provinces.

The destination effects are more variable. For provinces that deviate significantly from the average, it appears that there are different patterns, some of which correspond to those of internal migration and others that do not. For provinces that exhibit higher than average attraction, Beijing and Guangdong have known to be top migration destinations over the past three decades due to their high level of economic development. The rest of the provinces are mostly in Western China except for Hainan in the south. Among them, Xinjiang has also been a major migration destination due to its abundant natural and land resources and policy-led development (Fan 2005b; Liu, Stillwell et al. 2014). Neither Ningxia nor Hainan receive large volume of migrants but appear to have the top ten in-migration rate in the nation, which is related to the comparatively smaller population size (Liu, Stillwell et al. 2014). The rest of the Western provinces are more known for sending migrants instead of receiving. Another interesting fact about those provinces is that there is a distinctive pattern in terms of their urban migration rate. The three coastal provinces (Beijing, Guangdong and Hainan) all have a fairly high urban migration rate ranging from 0.91 in Hainan to 1.69 in Beijing (ranked 1st) and the Western provinces all have lower urban migration rate from 0.41 in Yunnan to 0.61 in Gansu. The above results suggest that even though the Western provinces are less developed and therefore not as active in stimulating urban-urban migration flows, these Western provinces exert higher than average attraction to urban migrants by offering other possibilities such as higher potential of income growth. This may be explained by the recent higher than average economic growth Western provinces have experienced due to the increased investment and preferential policies as part of China Western Development program sponsored by the central government.

In terms of provinces with below average attractiveness, except for Tianjin all are in less developed non-coastal areas with the majority located in central China. Even though Tianjin is a very developed municipality, its geographical proximity to Beijing perhaps explains why urban migration flows respond passively to its urban labour market (population) as Beijing is a more popular destination for urban migrants.

At the flow-pair level, an interesting fact is that among the flows that exhibit the largest and smallest residuals, the majority of them are between western provinces followed by those involving a western province as an origin or destination. Western provinces are the least developed area in China that have experienced fast policy-led development recently. Their economic growth could either act as a pull force due to increased economic opportunities or push force since the development would enable those who were not able to migrate to move to more developed areas now. Those facts might help contribute to their irregular or unpredictable urban migration flows from the more established migration streams. In addition, we also find high correlations between some of the pair-flows and most of those tend to be between Western provinces as well. For instance, the reciprocal migration flows between Yunnan and Xinjiang which are distantly apart are both significantly lower than predicted, while the bilateral migration flows between neighbouring provinces Sichuan and Tibet are both significantly higher than predicted. The latter result in a way mirrors international migration in less developed world. For example, most migrants in Africa and Southeast Asia tend to go to nearby countries instead of countries further away with more job opportunities. Only when countries become more developed, more people would start to engage in longer distanced migration.

5.6 Conclusions

This chapter has set out a method of multilevel modelling better to understand the patterns of interprovincial urban-urban migration in China. Whilst prior studies have greatly contributed to the understanding of migration flows, they wrongly treat migration flows between shared origins and/or destinations as independent events, leading to potentially inaccurate results, as well as overlooking substantively interesting correlations in the patterns of movement. In addition, little attention has been given to urban-urban migration flows. Our endeavour has overcome both research gaps, which are successfully considered in this chapter.

This study is the first to systematically analyse urban-urban migration in China, a phenomenon which is on the rise in developing countries that have been going through rapid urbanisation. Comparing to internal migration in China which is dominated by rural-urban migration, urban-urban migration is similar in the sense that it is also economically driven and larger population sizes at both origins and destinations help to contribute to the volume of the migration flows. Moreover, distance plays a sizeable deterring role on urban-urban migration, which may have important policy implications. On the one hand, for major cities in the western and inland areas, designing policies to manage distance's adverse effect can help reducing regional inequality by encouraging information, skill and capital transfer carried out by urban-urban migrants between them and the coastal cities. On the other hand, as China has been interested in initiating localised urbanisation by favouring developing small and medium sized cities since the 1980s (Han and Yan 1999; Chen, Liu et al. 2013), distance decay effect may reduce long distance urban-urban migration and as a result facilitate the local socio-economic development which will greatly benefit smaller cities.

The findings also suggest that development level is closely linked with urban-urban migration. For example, urbanisation level plays a key role in provinces' origin effect. The destination effect and pair-flow effects also indicate how the policy-driven growth in the least developed Western region have had great influence on urban-urban migration in China.

Chapter 6 Exploring the non-linearity of distance decay in interprovincial urban-urban migration flows

The last chapter discussed how flow dependencies in the traditional gravity model bias downwards the estimates of the standard errors for the regression coefficients. We responded to this problem by developing a multilevel gravity model to capture, measure and study these dependencies. Indeed, the model results have shown that allowing for flow dependencies greatly improves the prediction precision of regression coefficients by reducing the risk of type I errors of inference. Nevertheless, the proposed model of Chapter 5 treats the migratory distance between the origin and the destination as a linear term (i.e. the assumption of the linear relationship between log migration and log distance) in the model specification, which does not consider the situations wherein non-linear relationships may apply (Stouffer 1940; Zipf 1946; Stewart 1960; Sjaastad 1962; Olsson 1965; Jiang, Wang et al. 2013; Molho 2013).

This linear relationship between log migration and log distance comes from the log-log transformation of the original gravity model by taking logs of both sides of the original gravity models[90], which leads to a regression functional of log migration flow upon log distance in both Chapter 4 and 5. As indicated by earlier gravity models in this thesis, the marginal effect of log distance upon log migration remains constant [91]. That is, interprovincial migration drops a constant amount for a 1 km increase in between-province distance irrespective of between which provinces (or regions) this 1 km distance increase takes place. In reality, the marginal effect of log distance upon log migration can be non-constant. Rather it is a function of the distance at which the marginal effect is evaluated. For instance, an increase from 10 to 11 km will lead to a different drop in migration versus an increase from 1000 to 1001 km. In

[90] See Section 3.2 of Chapter 3, 4.4 of Chapter 4 and 5.3 of Chapter 5 for more details.

[91] As the province-pair level covariate vector only includes the log of provincial distance in earlier result chapters, the marginal effect of log distance upon log migration is constant and equals to the coefficient vector of the province-pair level covariate. See Sub-section 3.2.3 for more details.

theoretical terms, this linear relationship is thus based on assuming the continuity of provinces and regions, which underlies the constant rate of distance decay (here referring to the marginal effect of the log distance). In the regional migration system, however, continuity of provinces and regions usually matter due to distinctive provincial and regional socio-economic contexts. Indeed, Chinese provinces and regions are distinctive in their levels of development because of unbalanced regional development policies. On account of this, the multilevel gravity model proposed in the last chapter is restrictive to an extent by assuming a linear only relationship between migration and distance. Indeed, it would be naïve to believe that distance decay effect can be fully captured by a linear term (Olsson 1965; Wolpert 1966; Jiang, Wang et al. 2013).

This chapter therefore endeavours to explore the non-constant marginal effect of log distance upon log migration, by introducing a quadratic term of distance and including measurements of provincial and regional contiguity. Like in the last chapter, this model specification is tested with the interprovincial urban-urban migration flow in 2010 by using a subset of the data. This new approach has the potential to reveal the spatial distribution of migration flows in a more realistic way. This chapter consists of seven sections. The first section is the introduction, followed by the literature review on the measurement of distance decay. The third section describes the data, while the fourth section provides the model development process, with the subsequent fifth section presenting key results and findings. The sixth section further discusses and explains implications of the results, and the conclusion section summarises the knowledge advance in this chapter and thus maps out a research plan for future studies.

6.1 Introduction

Analysis and measurement of distance decay is of prime importance for studying migration flows (Stewart 1960; Morton 1977; Courgeau and Baccaini 1989; Cohen, Roig et al. 2008; Dennett and Wilson 2013; Thomas, Stillwell et al. 2015). As a general rule of migration studies, migrants are more likely to move to near rather than faraway places by following Tobler’s first law of geography (Tobler

1970), controlling for all other migratory factors (Shaw 1975). This corresponds to the typical and straightforward interpretation of a constant rate of distance decay under the gravitational law, as distance is inversely proportional to the rate of total migration (Courgeau and Baccaini 1989; Fan 2005b; Thomas, Stillwell et al. 2015).

Although this has undoubtedly helped our understanding of migration, the continuity of provinces and regions nevertheless usually matter in measuring distance decay, due to various geographical, political, cultural and institutional causes. For instance, Chinese provinces and regions are distinctive in their levels of development, which may lead to different rates of distance decay across provinces and regions. Other empirical examples of such geographic discontinuities in the global context include national identity building, governance, and ethnic segregation through political and institutional agents, which often result in changes in languages, social norms and economic activities across different administrative units (Posner 2004; Berger 2009).

Therefore, considering the continuity of both provinces and regions could improve the measurement of distance decay. The rationale is threefold. First, the continuity of provinces lies in neighbouring relationships between provinces. There are reasons to expect that the role of distance will differ for migration flows taking place between neighbouring and non-neighbouring provinces within a region. This is because neighbouring provinces are more likely to be close to each other and similar in migration and development policies, and social and cultural aspects such as dialects and social norms, which could further reduce the travel friction for migrants in making the movement. Secondly, regions may differ in their ability to attract and export migrants. Labour markets can be segregated and region-based as a result of unbalanced regional development policies, within which certain migration flows may benefit greatly whilst others are discouraged. Thirdly, provinces within the same region may differ in their ability to attract and export within- and between-region migrants. This is because differences in the returns on human capital may exist for provinces of the same region between being the home- and host-regional labour market, which can affect the ability of migrants to achieve income gains and thus induce further redistribution of migrants across regions. By situating the argument in the light of former studies, the following section will explain this threefold rationale in more detail.

6.2 Literature review

The threefold rationale of the non-linear distance decay lies in the combined impacts from the province neighbouring relationship and unbalanced regional development polices. The elaboration of these two effects is based on reviewing relevant literature in the following two sub-sections.

The literature so far offers important knowledge in understanding the definition of neighbouring relations, the association of migration and regional development policies in China, and the properties of dyads. This prior knowledge also enables this chapter to ask the following research questions about exploring the non-linearity of distance decay: (1) does introducing a quadratic term of distance improve the model fit? (2) how does the province-neighbouring relationship affect migration flows? (3) what effects do regions have upon migration flows?

To answer these research questions this chapter proposes an extended model to explore the non-linearity of distance, based on the multilevel gravity model of Chapter 5. This new model also aims to measure the effects of neighbouring-provinces and provincial-regions in interprovincial urban-urban migration, paying special attention to their interaction effects. The following section presents the relevant literature background.

6.2.1 Neighbouring relationships and migration

In geography, distance, interaction and connectivity are key concepts, whilst 'nearness' is the most important problem (Olsson 1965). 'Connectivity', or a neighbouring relationship, describes how places link with each other in space. If two (or more) spatial entities are linked to each other based on a connectivity criterion, they together are known as 'neighbours' of that connectivity standard (Lee 1968;

Jenks and Dempsey 2007). In general, neighbouring places tend to share similar spatial and socioeconomic characteristics (Crowder, South et al. 2006; Sun and Manson 2012; Jivraj, Brown et al. 2013). A neighbouring relationship can be an important factor in social sciences. For example, scholars find that democracy tends to appear and sustain in countries with neighbouring democratic states (Gleditsch and Ward 2006; Zhukov and Stewart 2013).

What constitutes a neighbour, or a choice of connectivity, has long been a fundamental concept in spatial analysis (Bavaud 1998; Anselin 2002; Kostov 2010) and a key research question for a wide range of social sciences (Gleditsch and Ward 2001; Buhaug and Gleditsch 2008; Zhukov and Stewart 2013). In migration studies, for example, there have been heated discussions regarding the definition and measurement of neighbouring relationships (Chun 2008; Dabbaghian, Jackson et al. 2010; Jivraj 2012; Sun and Manson 2012), particularly about calculations of sizes and contents of neighbours (Bailey and Livingston 2008; Sun and Manson 2012).

Although there are competing definitions of neighbours, most studies define neighbouring relationships by shared borders, and treat neighbours as a way to measure distance decay in social sciences (Morton 1977; Bavaud 1998; Gleditsch and Ward 2001; Anselin 2002; Buhaug and Gleditsch 2008). Neighbouring relationships have also been widely employed in measuring distance decay for migration studies around the world. For instance, studies have shown that socio-economic conditions of neighbouring communities influence migrants' assimilation (Jargowsky 2009), health and well-being (Crowder and South 2005; Buu, Mansour et al. 2007), as well as local community and city structure (Bailey and Livingston 2008; Sun and Manson 2012; Schlichting, Tuckel et al. 2015). Indeed, people's migratory behaviours in one geographical area could be influenced not only by compositions and characteristics of that area's population and socio-economic conditions, but also by the area's geographical contexts. In China's context, the attractiveness or emissivity of a boarder regional context that a province is situated in may have an important effect on people's migratory behaviours within that province.

Based on the valuable standard of neighbouring relationship measurement adopted by most studies, this chapter will utilise two concepts in the analysis to further refine distance decay: one is the neighbouring relationship of provinces, depending on whether provinces share borders or not; the other is the provincial region, defined as a group of provinces sharing the same regional development policy. The motivation behind this is to further explore the non-linearity of distance decay by considering the continuity of provinces and regions. First and foremost, provincial neighbouring relationship represents the existence of different spatial proximities and connectiveness, contributing to distinctive spatial patterns contained in the flows of neighbouring and non-neighbouring provinces (Chun 2008; Shen 2016b). Secondly, distance decay matters substantially to migration flows (Stewart 1960; Courgeau and Baccaini 1989; Molho 2013; Thomas, Stillwell et al. 2015). Indeed, distance is the defining element in distinguishing important genres of migration such as internal and international migration or more specifically in this study inter- and intra-provincial migration (Courgeau and Baccaini 1989; Molho 2013; Otoiu 2014).

Neighbouring-province in this thesis is defined as the first nearest neighbour (Keller, Gray et al. 1985; Arya, Mount et al. 1998), meaning that one migration flow takes place between a pair of neighbouring provinces if origin and destination provinces directly share common borders. The major reason to adopt this definition lies in that real networks of places connected by migration flows tend to have relatively short paths with strong links between any two nodes (Boccaletti, Latora et al. 2006; Chun 2008). A place primarily connects with its first nearest neighbours in terms of (social or geographical or both) distance through social networks and economic links (Moretti 1999; Newman and Park 2003; Boguná, Pastor-Satorras et al. 2004; Boccaletti, Latora et al. 2006; LeSage 2008). These social networks and economic links in turn have profound influences upon the regional migration system (Davis, D'Odorico et al. 2013). Studies have shown that distinguishing the first nearest neighbouring from non-neighbouring provinces could help to improve the measurement of distance decay, as the distribution of migration distances shows remarkable similarities with that of nearest-neighbour distances in stochastic migration models (Olsson 1965; Chun 2008; Levy 2010).

This definition has two major advantages. First, it discounts the size differences of provincial units (as the covariate of distance between provincial capitals already reflects the size of provincial units), ensuring that all the provincial units are treated equally as spatial entities. This is important, as neighbouring relationships and spatial links take place between provinces regardless of province sizes. Secondly, this definition can also ensure each spatial entity (or province) has the same number of enquiries and entries about its neighbouring relationships with all the other entities. The generated neighbouring matrix therefore fits well with the analysis framework based on individual migration flows (Zhukov and Stewart 2013).

The definition of the provincial-region is specific to the context of China. This is because the institutional drivers, namely unbalanced regional development policies, are widely considered as important in shaping and facilitating population relocation and redistribution in post-reform China (Fan 1997; Lin 2001; Villaverde, Maza et al. 2010; Pannell 2012). Indeed, unlike in other countries around the world, the state of China has greatly intervened in local development by implementing unbalanced regional policies, which has led to remarkable regional inequalities in China and has formed a close association with the massive internal migration across the country for the past few decades (Bao, Chang et al. 2002; Keidel 2009; Chen and Groenewold 2011). The following sub-section will explain regional development policies in China and the definition of provincial regions in more detail.

6.2.2 China’s regional development policies

A large body of studies has shown that regional development policies are closely associated with population movement within China. For instance, China’s massive internal population movement began in the late 1970s, exactly when the historical ‘reform and opening-up’ policy was first executed (Fan and Sun 2008; Ye, Wang et al. 2013). As such, it is of prime importance to understand the political context of China to rationalise the definition of provincial regions.

Recalling that regional policies have gone through five stages in general since 1949[92], policies of stage five are most recent and relevant to interprovincial migration in 2010, although all policies combined have greatly shaped the regional development landscape in China with cumulative effects over the years. Indeed, state interference in regional development is most clearly evidenced in the years post 2000 (Sun 2013). One major reason is the establishment of the National Development and Reform Commission (NDRC) in 2003, which has been overseeing the construction and implementation of regional development policies with increasing efficiency and strength (Sun 2013).

The three major economic region policies – China Western Development, Revitalise Northeast China, and Rise of Central China – have had profound impacts on the regional development landscape in China. Along with effects of earlier policies, implementation of these three policies has divided China into four distinct economic regions: the West, the Northeast, the Central and the East (Table 6.1). Especially post-2000, the state has spared no effort in alleviating regional inequality as evidenced by the three major economic region policies (Fan 1995; Démurger, Sachs et al. 2002; Fan, Kanbur et al. 2009). Nevertheless, there are still huge socio-economic inequalities between regions (Démurger, Sachs et al. 2002; Fleisher, Li et al. 2010; He, Bayrak et al. 2017): the East is the most developed, as it has enjoyed multiple and cumulative favourable policies since the reform (stage two onwards); the Central and Northeast regions are moderately developed, whilst the latter has been experiencing slow even stagnant growth recently; and the West is the most underdeveloped, though it was the second earliest region to enjoy favourable policies.

[92] See Sub-section 2.3.1.3 of Chapter 2 for more details.

Table 6.1 Timeline of China's economic region policies

Year	Policy	NDRC Department	Number of provincial units	Location
1978-now	A package of policies favouring coastal cities of the East region	None	10	Beijing, Tianjin, Hebei, Shandong, Jiangsu, Zhenjiang, Shanghai, Fujian, Guangdong and Hainan
2000	China Western Development	Department of Western Region Development	12	Chongqing, Shaanxi, Sichuan, Gansu, Qinghai, Yunnan, Guizhou, Guangxi, Inner Mongolia, Ningxia, Xinjiang, and Tibet
2003	Revitalise Northeast China	Department of Northeast Regional Revitalization	3	Heilongjiang, Jilin, and Liaoning
2004	Rise of Central China	Department of Regional Economy	6	Henan, Hubei, Hunan, Jiangxi, Anhui and Shanxi

A large body of research has shown how regional inequality is closely associated with population movement within China. But many divided the country into the East, the Central and the West three regions. For instance, Fan (2005a) analysed the interprovincial migration with 1990 and 2000 census data by dividing mainland China into eastern, central and western regions. The results indicated that the association between regional development and migration had strengthened, as had economic inequalities between provinces. Likewise, Bao, Shi et al. (2005) and Shen (2013) utilised the same three-region criterion as Fan (2005a) to investigate internal migration patterns through census data, arguing that migration growth from 1990 to 2000 was largely attributed to China's rapid and unbalanced economic development. Following the same region classification, Liu, Stillwell et al. (2014) found that the trend of concentrating migration destinations grew before 2000 but declined soon afterwards with evident changes in popular locations over the years. They argued that this mirrored the changing pattern of provincial economic disparities, demonstrating divergent trajectories in the 1980s and 1990s before the recent convergence since 2000.

With differing findings of regional divergence or convergence, the classical three-region division in former studies have undoubtedly improved our understanding of the regional development in China. The different findings, for instance, have described the regional development landscape from different perspectives and time spans. Particularly, scholars have widely acknowledged the sensitivity of the results towards different definitions of provincial regions (Lupton 2003; Jenks and Dempsey 2007; Zhukov and Stewart 2013). Studies have also highlighted the uniqueness of the Northeast region in particular: unlike all the other three regions, the Northeast used to be a heartland of powerful state-owned industrial sectors, which are rapidly shrinking and have caused a marked decline of economic growth (Lee 2007; Chovanec 2009). Contrasting with impressive growth rates of other regions, its marked regional decline is unique and thus the region is known as the Rust Belt in China (Lee 2007; Chovanec 2009).

On account of this, this chapter adopts a new four-region division - the West, the Northeast, the Central and the East. This definition is consistent with both the economic-political and geographical contexts of China, which is particularly helpful to understand the unique role of the Northeast in the regional migration system. A migration flow is defined as originating from a certain region if its origin province belongs to this region, and as ending in a certain region if its destination province is from this region. In other words, one migration flow has two regional attributes, the origin and the destination regions.

6.3 Data

The data used in this chapter are the same as those in the last chapter, except for the newly-added covariates of the neighbouring provinces and the provincial regions whose definition and measurement will be detailed here. Systematic explanations about all the other variables can be found in earlier chapters[93].

[93] See Section 5.1 of Chapter 5.

Table 6.2 The number of migrants contained in flows of neighbouring provinces (unit: persons)

Neighbouring provinces	Mean	S.D.	Freq.
No (0)	109,042	258,691	790
Yes (1)	346,439	644,246	140
Total	144,780	355,179	930

Two neighbouring provinces are linked by a migration flow[94]: between these two provinces, one is the origin and the other the destination. Provincial neighbours are defined by whether two provinces share a boundary. This is represented by a binary variable, where '1' represents a pair of neighbouring origin and destination provinces and '0' represents non-neighbouring provincial pairs. The variable of neighbouring provinces has 930 observations, among which 140 are coded as '1'. In other words, 140 out of all the 930 urban-urban migration flows take place between neighbouring provinces (Table 6.2). Migration flows of provincial neighbours have a larger number of migrants on average (346,439 persons) and they vary more (S.D.=644,246 persons), compared with those of provincial non-neighbours (Mean=109,042 persons; S.D.= 258,691 persons). This is related to the fact that neighbouring provinces tend to be near to each other, so they are subject to the adverse effect of distance (i.e. the growing cost of moving for migration with increasing distances) to a lesser extent. Additionally, the neighbouring relationship also represents the connectivity between the origin and the destination provinces. In other words, this enables the exploration of the continuity of the between-province boundaries for migration flows. Moreover, the covariate of the provincial capital distance is a proxy measurement of the actual between-province migratory distances, which are unavailable in the datasets. However, provinces vary in sizes, therefore neighbouring does not necessarily guarantee a short between-province migratory distance. Despite this, these reasons lend support to the belief that the provincial neighbours and non-

[94] See more explanation in Sub-section 6.2.1.

neighbours are two distinct conceptual groups, which might require different treatments in the model specification.

As shown in Figure 6.1, this chapter adheres to NDRC's definition of economic regions. To elaborate, the East and the West each consist of 10 and 12 provinces, whilst the Central region contains 6 provinces, with 3 provinces (Heilongjiang, Jilin and Liaoning) forming the Northeast region as shown in Table 6.1[95]. This four-region division contrasts with the classical three-region division criterion in former studies (Bao, Shi et al. 2005; Fan 2005a; Shen 2013), which divides China into the East, West and Central regions by treating Liaoning as an eastern-region province and Heilongjiang and Jilin as central-region ones. This chapter will primarily focus on using the four-region division criterion based on NDRC's definition in Model 1 to 3, as the research interest of this chapter is to explore the role of the Northeast region. Model 3a will robust-check the results by using the three-region criterion.

[95] Table 6.1 from Sub-section 6.2.2 provides more details about the composition of provinces for each region.

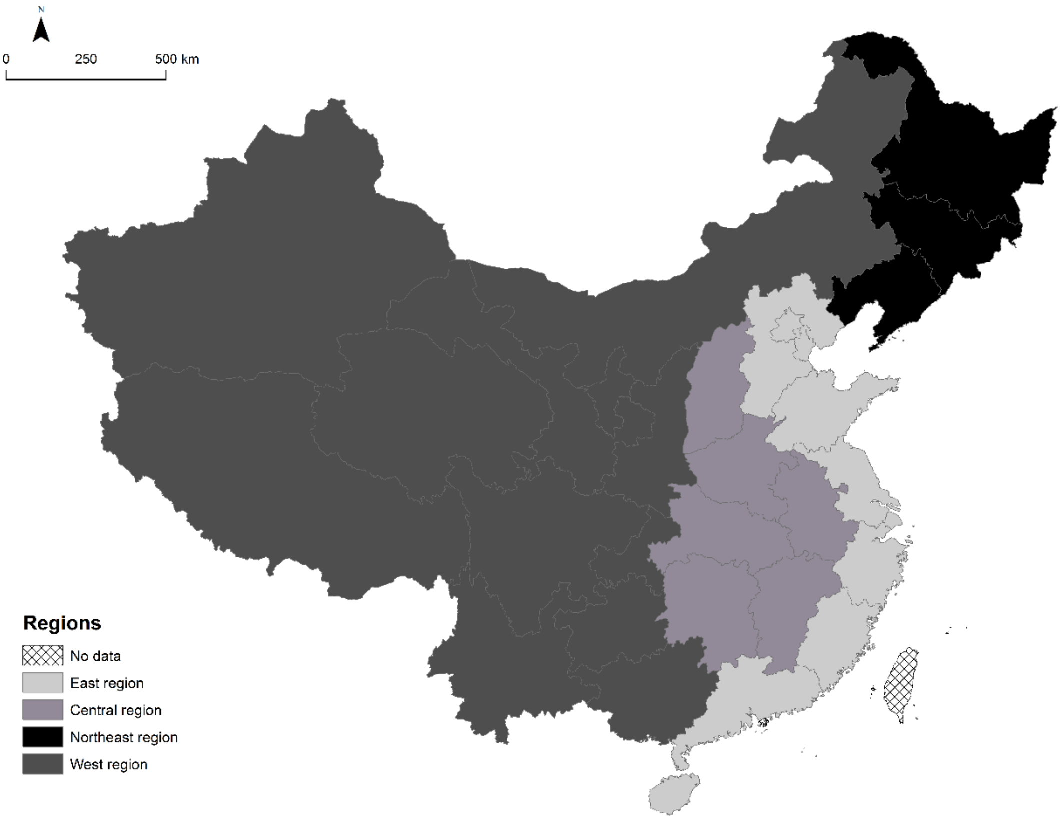

Figure 6.1 Definition of economic regions in China

Table 6.3 is effectively a two-way cross-tabulation presenting the row and column percentages, which are calculated by dividing the number of migrants in each flow with the total number of migrants in the in- or out-flows for each region respectively. Like former chapters, within-province migration is excluded from all the calculations. Percentages of regional flows in Table 6.3 show the distribution of migration across regions, highlighting regional emissivity differences and destination preferences among urban-urban migrants. For instance, the share of in-migration within the East stands at 27.17%, representing that 27.17% of all the in-migrants ending at the East come from provinces of the East. This high portion shows that within-region flow is relatively strong for the East. The share of out-migration is 68.21% within the East, meaning that 68.21% of all the out-migrants originating from the East also ends at the East. This further manifests the predominate preference of within-region relocation among out-migrants of the East. In contrast to this 68.21% within-region flow proportion of the East, the share of out-migration stands at 5.52% between the East and the Northeast, representing that only 5.52% of all the out-migrants originating from the East end at the Northeast provinces. This demonstrates how unpopular the Northeast is for the out-migrants of the East. The share of in-migration between the East and the Northeast is 13.76%, meaning that only 13.76% of all the in-migrants ending at the East originate from provinces of the Northeast. This Northeast-East flow has the smallest share among all the in-migration flows for the East, indicating that the Northeast is dwarfed by other regions (including the East itself) as the migration source for the East.

Table 6.3 The shares of in- and out-migration for each region (%)

		East	Northeast	Central	West	Share Total
East	In-migration	27.17	13.76	38.10	20.97	100.00
	Out-migration	68.21	5.52	11.94	14.33	100.00
Northeast	In-migration	25.36	46.33	12.31	16.00	100.00
	Out-migration	67.34	19.66	3.85	9.16	100.00
Central	In-migration	42.67	7.05	29.23	21.04	100.00
	Out-migration	79.10	2.22	6.76	11.92	100.00
West	In-migration	23.94	7.84	24.08	44.13	100.00
	Out-migration	59.53	3.94	6.66	29.88	100.00
Total Flow	In-migration	69.61	6.03	7.76	16.60	100.00
	Out-migration	27.73	14.22	33.53	24.52	100.00

Overall, the imbalance of net flows for a region is given by the gap between the in- and out-migration shares of the total flow for each region (Table 6.3). Specifically, the East receives the most migrants in total (share 69.61%; 937,236 persons), whilst the Central region sends out the most by comparison (share 33.53%; 451,457 persons). Overall, however, only the East has a net in-flow of migrants (69.61%-27.73%=41.88%; 563,895 persons), whilst the remaining three regions all suffer from net loss of population to different extents.

Table 6.4 The mean and S.D. of the number of migrants per flow for each region (original data; unit: persons)

			Destination region				
			East	Northeast	Central	West	Total
Origin region	East	Mean	282,971	68,667	74,292	44,599	124,456
		S.D.	491,511	63,613	62,373	49,213	291,502
		Frequency	90	30	60	120	300
	Northeast	Mean	429,753	627,183	40,917	48,703	212,728
		S.D.	458,492	538,939	14,420	65,084	361,704
		Frequency	30	6	18	36	90
	Central	Mean	595,148	55,556	101,787	74,767	250,809
		S.D.	895,370	37,931	70,322	76,870	572,093
		Frequency	60	18	30	72	180
	West	Mean	163,789	36,111	30,528	74,730	91,714
		S.D.	337,544	75,483	27,297	154,564	223,810
		Frequency	120	36	72	132	360
	Total	Mean	312,412	90,257	58,031	62,091	144,780
		S.D.	567,790	202,913	56,054	106,363	355,179
		Frequency	300	90	180	360	930

Note: Due to each region containing different numbers of provinces, the number of migration flows (the frequency) varies considerably across regions. There are 12, 10, 6 and 3 provinces in the West, East, Central and Northeast region respectively.

Dividing the total in- and out-migration population for each region by the total number of flows (the frequency) shows a different mean number of migrants per flow for each region (Table 6.4). For instance, the first cell of 282,971 in Table 6.4 shows that there were about 282,971 migrants per flow on average for the 90 within-region flows[96]. Judging by the means and variances in Table 6.4, the four regions are distinct from each other, which is important to acknowledge in the following model specification. Specifically, migration flows of the Central and the Northeast contain larger number of out-migrants per flow on average (250,809 and 212,728 persons in Table 6.4) and have larger standard deviations (572,093 and 361,704), in comparison with those of the East and West regions (Table 6.4). Echoing the earlier finding about the total net flows, the East differs clearly from the other three regions by being the most attractive region and receiving the largest number of in-migrants per flow on average

[96] Recalling the 10% sampling rate explained in Section 3.1 of Chapter 3.

(312,412 persons). By contrast, the Central region attracts the lowest number of in-migrants per flow on average (58,031 persons).

All regions tend to have higher average number of migrants per flow within a region (i.e., within-region flows as shown the diagonal cells in Table 6.4) compared with between-region flows that either only originate from or end at that specific region. Specifically, the within-region flow of the East (282,971 persons) has the largest number of out-migrants per flow on average, compared with all the other out-flows of the East. By contrast, the within-region flow of the Central region (101,787 persons) has the largest average number of in-migrants, compared with all the other in-flows of the Central region. The within-region flow of the Northeast (627,183 persons), however, has the largest average number of migrants per flow within the region for both the in- and out-migration flows, compared with all the other in- and out-flows of the Northeast respectively. Meanwhile, the West has the second largest average number of migrants per flow (74,730 persons) within the region for both in- and out-migration flows, compared with all the other in- and out-flows of the West respectively. All these show that within-region flows might be empirically different from between-region ones, which will be further explored in the modelling process.

6.4 Methods

In this section, the gravity model of migration from Chapter 5 is treated as the base model (Model 0), which is further modified by relaxing the linear function of the natural logarithm distance as Model 1. And then, this chapter presents two increasingly realistic specifications of distance decay: Model 2, the neighbouring-province model, extends Model 1 to capture the effect of neighbouring provinces; and Model 3, the model incorporates both the neighbouring-province and provincial-region effects along with the interaction effects between the neighbouring-province and provincial-region covariates. Model 1 allows the exploration of the non-linear function of distance decay, and on top of that Model 2 further considers the contiguity of provinces. Apart from the non-linearity of distance decay and the continuity of provinces, Model 3 additionally investigates whether regional differences matter for migration flows so as to answer the three aforementioned research questions.

Furthermore, the appendix shows the development and results of three additional models. Model 2a is the first intermediate model that investigates only the provincial-region effects. Model 2b is the second intermediate model incorporates both the neighbouring-province and provincial-region effects but excludes the interaction effects between the neighbouring-province and provincial-region covariates. Model 3a is the robust-check model, which is a modified version of Model 3 by adopting the traditional three-region definition in place of the four-region definition. This section then shows how Model 3 measures all distance decay effects of the research interest. Like the former chapter, this chapter conducts the estimation and calculation of all the three models with runmlwin in Stata SE 14 (Leckie and Charlton 2013).

6.4.1. Model 1: Relaxing the linear function in the multilevel gravity model of migration

If the relationship between the log migration flow $\ln(m_{ij})$ and the log of the distance $\ln(d_{ij})$ is not linear, one approach to capture this non-linearity is to add polynomial terms. Therefore, Model 1 relaxes the linear function between log migration and log distance by adding a quadratic term in the province pair-level covariate vector $\mathbf{x}_{3ij}$ of Equation (5.2)[97], specified as

$$\mathbf{x}_{3ij} = \left[\ln(d_{ij}),\ (\ln(d_{ij}))^2\right] \tag{6.1},$$

where $(\ln(d_{ij}))^2$ denotes the quadratic term of the log distance, and the coefficient vector $\boldsymbol{\beta}_3$ also changes accordingly. By doing this, it is now possible to explore the non-linear relationship between log migration and log distance. In prior estimations, the marginal effect of log distance upon log migration is hold constant due to the linearisation process. In equation 6.1, the marginal effect of log

[97] See Sub-section 5.3.3 for more details.

distance upon log migration becomes $\boldsymbol{\beta}_3$*[1, 2ln$\left(d_{ij}\right)$)], which is now dependent on the value of log distance and not constant.

6.4.2. Model 2: The neighbouring-province

Based on the specification of Model 1, Model 2 can further incorporate the neighbouring-province covariate n_{ij} into the province pair-level covariate vector $\mathbf{x}_{3ij}$ of Equation (6.1), which can be written as

$$\mathbf{x}_{3ij} = \left[\ln\left(d_{ij}\right),\ (\ln\left(d_{ij}\right))^2, n_{ij}\right] \qquad (6.2),$$

where n_{ij} is a binary covariate equal to n_{ji} and is defined as

$$n_{ij} = \begin{cases} 0, if\ i\ and\ j\ are\ not\ neighbouring\ provinces; \\ 1, if\ i\ and\ j\ are\ neighbouring\ provinces. \end{cases} \qquad (6.3).$$

The coefficient vector $\boldsymbol{\beta}_3$ of $\mathbf{x}_{3ij}$ also further expands accordingly. Particularly, including the non-log term of the discrete covariate n_{ij} in the Equation (6.2) means that its coefficient (denoted as β_n) has an interpretation (Leamer 2012) different from that of log terms: that is, n_{ij} only has two values '0' and '1', and changing from '0' to '1' is associated with a $100 * (\exp(\beta_n) - 1)$% increase in the dependent variable m_{ij}.

6.4.3. Model 3: The neighbouring-province and provincial-region

Based on the establishment of Model 2, Model 3 further incorporates the provincial-region terms in the formulation, including covariates for within-region flows, the origin region as well as the destination region and their interaction terms. r_{ij} denotes the within-region flows[98], which is a province-pair level binary covariate and is defined as

$$r_{ij} = \begin{cases} 0, if\ i\ and\ j\ are\ not\ in\ the\ same\ region; \\ 1, if\ i\ and\ j\ are\ in\ the\ same\ region. \end{cases} \qquad (6.4).$$

Three dummy variables r_{1i}, r_{2i}, and r_{3i} specify the four origin regions and another three dummy variables r_{1j}, r_{2j}, and r_{3j} specify the four destination regions. In each case the reference category is set to be the East region, as it is the most developed region and hypothetically should be the most active in sending and receiving migrants according to the neo-classical economic theory of migration (De Haas 2010a, 2010b). The subscripts 1, 2 and 3 denote the remaining Northeast, Central and West regions respectively. Thus, the three origin region dummies measure how much higher out-migration is (on the log scale) in each of these three regions compared to the East. Similarly, the three destination region dummies measure how much higher in-migration is in each of these three regions compared to the East.

Consequently, the current equation for Model 3 can be written as:

$$\mathbf{x}'_{1i} = [\ln(p_{ui}), \ln(I_{ui}), r_{1i}, r_{2i}, r_{3i}]$$

$$\mathbf{x}'_{2j} = \left[\ln(p_{uj}), \ln(I_{uj}), r_{1j}, r_{2j}, r_{3j}\right]$$

$$\mathbf{x}_{3ij} = \left[\ln(d_{ij}), (\ln(d_{ij}))^2, n_{ij}, r_{ij}\right] \qquad (6.5),$$

Their corresponding coefficient vectors $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ further expand to capture the coefficients of the newly added covariates. Similar to the coefficient of n_{ij}, coefficients of

[98] Similar to n_{ij}, r_{ij} is also exchangeable, which means that r_{ij} is equal to r_{ji}.

the new dummy variables have the same interpretation of non-log discrete covariates (Leamer 2012). Namely, these new dummy variables all have their limited values, within the range of which a 1 unit increase in a covariate is associated with a specific percentage increase in the dependent variable m_{ij}; and that specific percentage change is calculated as $100 * (\exp(\beta) - 1)\%$ for the new value of that covariate (β is the generic regression coefficient for any dummy covariate of r_{ij}, r_{1i}, r_{2i}, r_{3i}, r_{1j}, r_{2j}, and r_{3j}).

However, there is little reason to believe that the effect of the neighbouring province does not depend on the region. For instance, a pair of neighbouring provinces in the East are not likely to share similar number of migrants per flow on average between them with those in the other three regions (Table 6.4). Therefore, the final equation for Model 3 can be written as follows by further including the interaction terms between the neighbouring-province and provincial-region variables.

$$y_{ij} = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{x}_j'\boldsymbol{\beta}_2 + \mathbf{x}_{ij}'\boldsymbol{\beta}_3 + o_i + d_j + e_{ij}$$

$$\begin{pmatrix} o_i \\ d_i \end{pmatrix} \sim N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_o^2 & \\ \sigma_{od} & \sigma_d^2 \end{pmatrix}\right\}$$

$$\begin{pmatrix} e_{ij} \\ e_{ji} \end{pmatrix} \sim N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & \\ \sigma_{ee} & \sigma_e^2 \end{pmatrix}\right\}$$

$$\mathbf{x}_{1i}' = [\ln(p_{ui}), \ln(I_{ui}), r_{1i}, r_{2i}, r_{3i}]$$

$$\mathbf{x}_{2j}' = \left[\ln(p_{uj}), \ln(I_{uj}), r_{1j}, r_{2j}, r_{3j}\right]$$

$$\mathbf{x}_{3ij} = \left[\ln(d_{ij}), (\ln(d_{ij}))^2, n_{ij}, r_{ij}, n_{ij}r_{1i}, n_{ij}r_{2i}, n_{ij}r_{3i}, n_{ij}r_{1j}, n_{ij}r_{2j}, n_{ij}r_{3j}\right] \qquad (6.6).$$

Here, $n_{ij}r_{1i}, n_{ij}r_{2i}$ and $n_{ij}r_{3i}$ are the interaction terms of the neighbouring-province and origin-region covariates, and $n_{ij}r_{1j}, n_{ij}r_{2j}$ and $n_{ij}r_{3j}$ are the interaction terms of the neighbouring-province and destination-region covariates[99].

In addition, three more models are developed as simplified versions of Model 3 to fully answer the research questions [100] . Specifically, Model 2a only incorporates the 7 regional dummies $r_{1i}, r_{2i}, r_{3i}, r_{1j}, r_{2j}, r_{3j}$ and r_{ij} on the basis of Model 0, so as to examine how regional effects alone affect interprovincial urban-urban migration flows. Model 2b additionally incorporates the neighbouring-province covariate n_{ij} in the equation of Model 2a. Note that Model 2b represented by Equation (6.6), differs from Model 3 by excluding the interaction effects between the neighbouring-province and provincial-region covariates. In other words, comparing Model 2b and 3 could help better understand the interaction terms. Nevertheless, it is necessary to robust-check three-region definition with an additional Model 3a, which incorporates only two dummy variables r_{2i} and r_{3i} for the three origin regions and another two dummy variables r_{2j} and r_{3j} for the three destination regions. The reference category remains as the East region, and the subscripts 2 and 3 denote the remaining Central and West regions respectively in place of the four-region specification in Model 3.

6.5 Results

This section will demonstrate the need to relax the linear assumption of distance with Model 0 (the final model in Chapter 4) and 1 (further including a quadratic term of log distance based on Model 0) in the first Sub-section. It will then compare the results from Model 1 to 3 (Model 2 extends Model 1 to capture the effect of neighbouring provinces whilst Model 3 incorporates both the neighbouring-province and provincial-region effects along with their interaction effects) in the second Sub-section. It

[99] Note that there is no need to further include an interaction term of the neighbouring-province and within-region variables $n_{ij}r_{ij}$, as these former 6 interactions terms already capture its effects.
[100] Further details about their equations can be found in the Appendix.

will subsequently explain why Model 3 is of best-fit. After that, this section will focus on explaining the fixed and random effects of Model 3 in the third Sub-section, based on comparison with those of Model 1 and 2. Some important findings of Model 3 needed to highlight are: (1) The effects of the neighbouring province; (2) The within-region effect of flows and the origin and destination effects of regions; and (3) The interaction effect of the neighbouring-province and provincial-region effects.

6.5.1 Relaxing the linear assumption of distance

Results of Model 1 show that the association between log migration and log distance is only statistically significant for the quadratic term[101], and Model 1 is a statistically significant improvement on Model 0 (the final model proposed in Chapter 5) (Likelihood-ratio test, $\chi_1^2 = 7.23, p < 0.01$). This indicates that adding the quadratic term improves the overall model fit, confirming, at least statistically (as opposed to substantively), the need to relax the linear assumption between log migration and log distance. Figure 6.2 presents the relationship between log migration and log distance with linear and quadratic fitted lines from Model 0 and 1 respectively, by holding all the other covariates at their means. As shown by Figure 6.2, the two fitted lines both demonstrate the negative association of log distance and log migration, but they appear very similar substantively.

[101] The result is sensitive to how log distance is centred. As this thesis does not centre covariates, readers are encouraged to not overly interpret the non-significance of the linear term here. What is important here therefore is primarily the overall test comparing model 1 and 0.

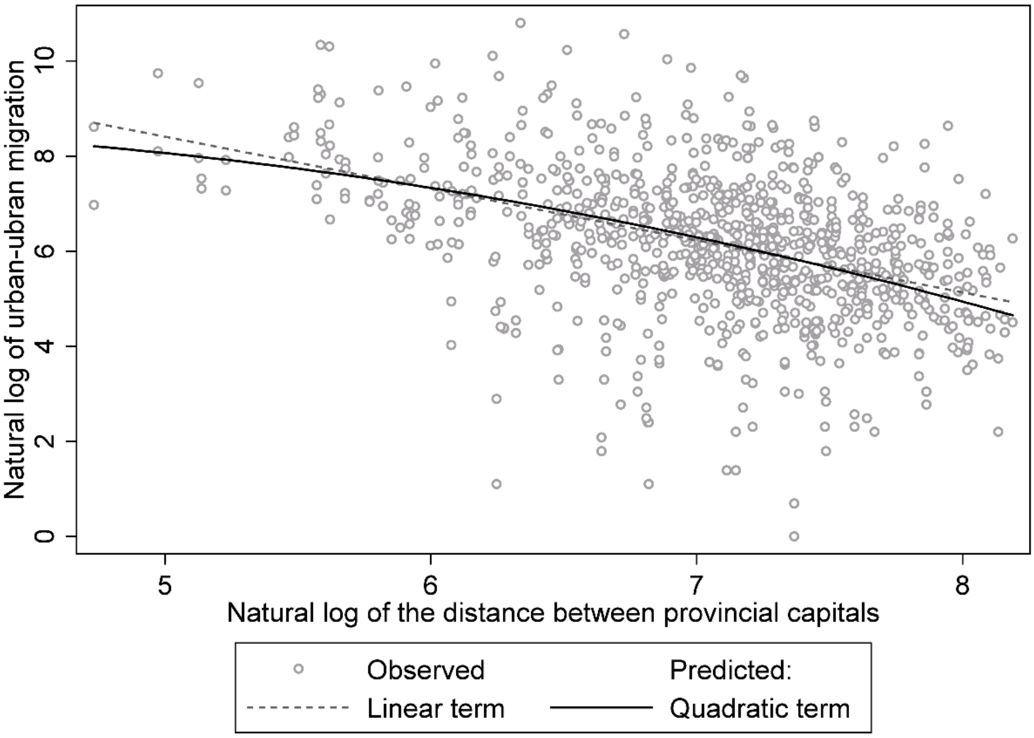

Figure 6.2 Comparison of fitted lines of Model 0 and 1 (Units: km and people in thousands; the points relate to the 930 province pairs)

Furthermore, including both linear and quadratic terms in the equation also better reflects the reality, which inherently suggests a peak value of the number of migrants for a range of migratory distances. This convex relationship is as expected and in line with the theory: there is a turning point of the log distance, before which the marginal effect of distance decay is positive but after which it turns negative (De Jong and Fawcett 1981; De Haas 2010a; Molho 2013; Otoiu 2014). Many empirical studies have also confirmed this pattern: there exists a balancing point where the growth of travelling cost is equal to the increase of destination choices that represent the likelihood of migration return, with the rising migratory distance acting as the search radius (De Jong and Fawcett 1981; Evans 1990; De Haas 2010a; Meeus 2012; Molho 2013; Otoiu 2014). As Model 1 stands, the estimated turning point is 23 km (3.132 on the log scale, when exp (3.132) = 23). This is far below the minimum distance between provincial capitals (114 km) and may make the curvilinear finding to be not substantively important at first sight. Nevertheless, the model-indicated non-constant marginal effect of log distance upon log migration is still meaningful, in that it still indicates that the relationship between log migration and log distance can be complex in other ways. For instance, the non-linear distance decay can be very much subject to the impacts of neighbouring provinces and provincial regions[102]. This initial interpretation of insignificant curvilinear finding therefore does not consider the neighbouring relationship between provinces and the regional effect. Indeed, the neighbouring relationship between provinces and the regional effect will alter this initial interpretation and will be further explained in the following. The non-constant marginal effect of log distance is therefore deserves further exploration and is worth pursuing.

Although non-linear terms allow for a large number of possible polynomial functions, the coefficient of the cubic term is not significant at 95% level of confidence, thus adding a cubic term does not significantly improve the overall model fit (Likelihood test, $\chi_1^2 = 3.31$, $p = 0.07$). Therefore, the specification of the polynomial term in Model 1 does not include a cubic or higher exponent term.

6.5.2 Overall model fit comparison

This sub-section will compare the models regarding the overall model fit, by primarily focusing on Model 1, 2 and 3. In doing so, the model comparison can be carried out thoroughly and systematically, which facilitates the choice of the best-fit model.

Table 6.5 shows the results from models 0, 1, 2 and 3. All four models contain some common covariates, including the natural log of the origin and the destination populations and incomes as well as the natural log of the distance between provincial capitals[103]. Recall that Model 2 extends Model 1 by including the neighbouring-province effect, whose coefficient is 0.623 and significant at the 99% confidence interval in Model 2. It shows that on average the volume of flows is 87% larger for provincial neighbours than that of non-neighbours (exp (0.623)-1 = 0.865). Wald tests confirm that Model 3 is

[102] See Section 6.2 of Chapter 6 for more details.
[103] Note that the focus of this chapter is the non-linearity of distance decay so the primary investigation is on the quadratic term of the log distance.

significantly preferred to Model 2 and 1, by showing the significant within-region effect ($\chi_1^2 = 19.14$, $p < 0.001$), the destination region ($\chi_3^2 = 15.02$, $p < 0.01$), and the interaction between neighbouring province and origin region ($\chi_3^2 = 16.17$, $p < 0.01$) and that of neighbouring province and destination region ($\chi_3^2 = 23.28$, $p < 0.001$) in Model 3. Moving from Model 1 to 2, the neighbouring-province covariate alone helps to explain 9% of the total residual variance ((0.698-0.636) / 0.698 * 100% = 9%), whilst covariates of both neighbouring-province and provincial-region altogether account for 17% of the total residual variance moving from Model 2 to 3 ((0.636-0.528) / 0.636 * 100% = 17%).

Table 6.5 Results from the four models

	Model 0		Model 1		Model 2		Model 3	
Variables	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Fixed part								
Constant	7.410***	1.503	0.760	2.905	-5.235	2.869	2.370	3.325
Log of origin urban population	0.909***	0.074	0.907***	0.077	0.901***	0.073	0.894***	0.084
Log of destination urban population	0.620***	0.105	0.618***	0.110	0.611***	0.105	0.767***	0.098
Log of origin urban income	-1.064***	0.283	-1.086***	0.295	-0.952**	0.280	-1.096*	0.436
Log of destination urban income	2.108***	0.401	2.086***	0.420	2.220***	0.400	1.374**	0.508
Log of distance	-1.091***	0.051	0.971	0.747	2.228**	0.734	0.987	0.757
Square of log of distance			-0.155**	0.056	-0.229***	0.054	-0.146**	0.056
Neighbours (reference: Non-neighbours)					0.623***	0.088	0.241	0.178
Within-region (reference: Between-region)							-0.262***	0.060
Origin region (reference: East)								
Northeast							0.007	0.283
Central							-0.200	0.237
West							-0.182	0.222
Neighbouring province*Origin region (reference: Neighbours*Origin East)								
Neighbours*Northeast							0.238	0.221
Neighbours*Central							0.097	0.139
Neighbours*West							0.491***	0.133
Destination region (reference: East)								
Northeast							-0.605	0.329
Central							-0.946**	0.276
West							-0.249	0.258
Neighbouring province*Destination region (reference: Neighbours* Destination East)								
Neighbours*Northeast							0.351	0.221

Neighbours*Central							-0.254	0.139
Neighbours*West							0.349**	0.133
Random part								
Origin province variance	0.104***	0.030	0.114***	0.032	0.102***	0.029	0.101***	0.029
Destination province variance	0.221***	0.060	0.244***	0.065	0.221***	0.059	0.141***	0.039
Individual flow variance	0.348***	0.021	0.340***	0.020	0.313***	0.018	0.286***	0.016
Origin-destination correlation	0.105	0.193	0.194	0.186	0.098	0.192	0.116	0.193
Flow-pair correlation	0.719***	0.023	0.712***	0.024	0.687***	0.025	0.675***	0.026
Deviance	1503.3		1496.1		1448.7		1381.0	
Total residual	0.673		0.698		0.636		0.528	

Note: Response variable is the log migration flow (in 1000s). *** denotes p < 0.001, ** denotes p < 0.01, and * denotes p < 0.05.

Furthermore, additional likelihood ratio tests with models in the Appendix further confirm that Model 3 is indeed the model of the best fit. For instance, Model 2a, which includes the quadratic log distance term but only the provincial-region variables, is significantly preferred to Model 1 ($\chi_7^2 = 23.80, p < 0.01$), whist Model 3 is significantly preferred to Model 2a ($\chi_7^2 = 91.26, p < 0.001$). This shows that adding both neighbouring-province and provincial-region variables raises the overall model fit most, although adding provincial-region variables alone could also raise the overall model fit to some extent. Moreover, Model 3 is significantly preferred to Model 2b ($\chi_6^2 = 36.66, p < 0.001$), evidencing the need to consider the interaction effect between the neighbouring-province and provincial-region covariates.

In order to examine whether results of Model 3 are robust to the choice of the traditional three-region definition, Model 3 is refitted with a three-category definition of region as Model 3a. Model 3 has lower deviance than Model 3a ($\chi_4^2 = 6.92$), which includes both neighbouring-province and provincial-region variables but adopts the three-region definition. Although the likelihood ratio test is not significant ($p > 0.05$) enough to confidently reject the hypothesis that both three- and four-region division criteria apply to China at a reasonable significance level, there is no doubt that the four-region definition is a better way to explore the role of the Northeast in the national development landscape. For instance, Wald tests of Model 3 show that the destination effect of the Northeast is significantly different from that of the Central region for flows between neighbouring provinces ($\chi_1^2 = 7.45, p < 0.01$).

6.5.3 Fixed and random effects of Model 3

In all four models, most of the estimated coefficients of common covariates have expected signs, and statistically significant (Table 6.5). Moreover, coefficients of common covariates except for those of the log distance are similar across models, meaning that their interpretations are similar and have been detailed in Chapter 4. Therefore, this sub-section will only briefly summarise the interpretation of most common covariates here, as the major focus is to investigate the non-linearity of distance decay from

Model 1 to 3. For instance, all four models show that the larger the origin urban population, the greater the flow: a 10% increase in origin urban population is associated with an approximate 9% increase in out-migration across all the models ($100 * (1.10^{0.9} - 1)\% = 10\%$), all else equal. This positive association with the volume of flows also holds true for destination urban population, but comparatively weaker than that of origin urban population across all the models. This is in line with findings of Chapter 3 and 4, meaning that the larger the origin or destination urban population the greater the flows. Lower origin urban income is related to more out-migrants, whilst higher urban income of the destination attracts stronger in-migration flows. The log distance, however, has positive coefficients across all three models, but is insignificant in both Model 1 and 3. By contrast, the negative association between the quadratic term of the log distance and the volume of flows is pronounced throughout all three models from Model 1 to 3. On the whole, therefore, this shows that the relationship between migration and distance is clearly non-linear, again re-asserting the need to relax the assumption of linearity.

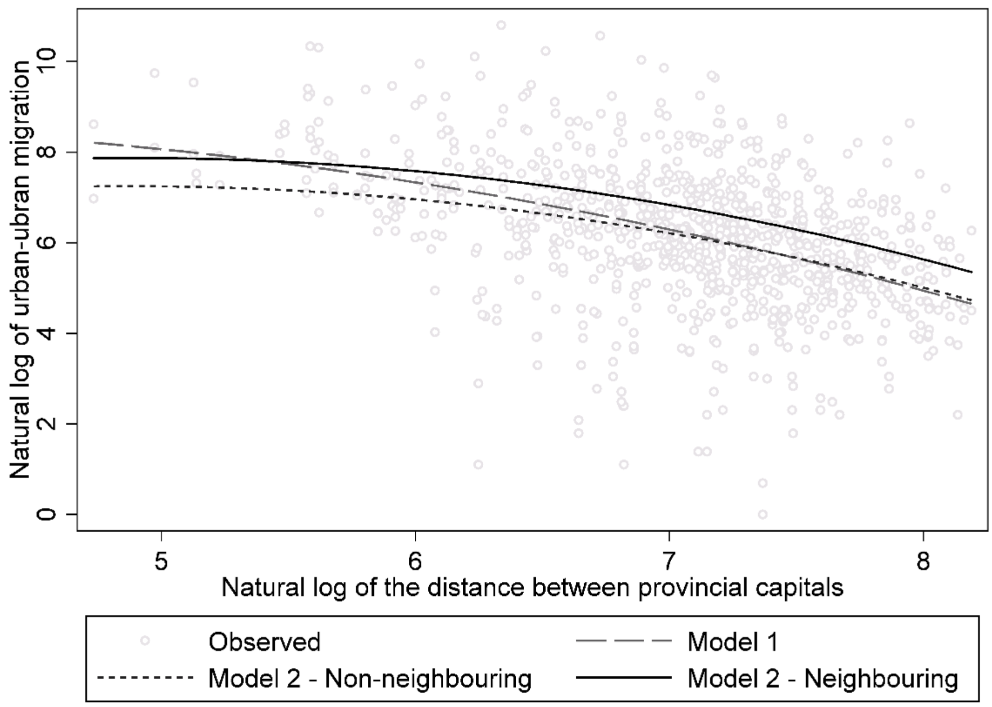

Figure 6.3 Comparison of fitted lines of Model 1 and 2

Moving from Model 1 to Model 2, the coefficient of log distance increases noticeably and becomes significant, whilst constant and the coefficient of the squared log distance decreases (Table 6.5). This indicates that the neighbouring-province covariate has brought some notable changes in Model 2, with effect of distance decay evidenced sensitive to the provincial neighbouring relationship. In Model 1 and 2, for instance, the relationship between log migration and log distance is not linear so the marginal effect of the distance on the migration is not constant, as shown in Figure 6.3 by holding all the other covariates at their means. To be more specific, 23 km and 30 km are the theoretical turning points in Model 1 and 2 respectively where distance has the smallest adverse effect upon the volume of the migration flow (when exp (-0.971/(-0.155*2)) = 23 and exp (-2.228/(-0.229*2)) = 30). However, all Chinese provincial capitals are located at least 114 km away from each other in reality, which implies a non-linear negative relationship between log distance and log migration within the observed distance span in both Model 1 and 2 (Figure 6.3). Most importantly, the marginal effect of distance decay decreases moving from Model 1 to 2, as shown in Figure 6.3. To be more specific, the fitted quadratic line of Model 1 cuts through the two-parallel fitted quadratic lines of the neighbouring and non-neighbouring flows in Model 2 from the top to the bottom within the range of observed log distance. Simply put, an increase in distance is associated with less decrease in migration when all the other covariates are held fixed moving from Model 1 to 2.

Furthermore, the neighbouring-province effect is captured as both significant and sizeable in Model 2 (Table 6.5). In fact, on average the volume of flows is 87% larger for provincial neighbours than that of non-neighbours as shown in Figure 6.3 (exp (0.623)-1 = 0.865[104]). Moreover, this captured neighbouring effect is accompanied by the decreased marginal effect of distance decay moving from Model 1 to 2 (Figure 6.3). Therefore, it illustrates that Model 1 incorrectly overestimates the marginal effect of distance decay, by not separating the effect of neighbouring-province from that of distance. Indeed, the neighbouring-province relationship could balance out some distance decay, due to both the nearness and connectivity between neighbouring provinces, as shown by the fitted quadratic line of migration flows between neighbouring provinces. In summary, Model 1 is inadequate for modelling distance decay or the neighbouring-province effect of migration flows.

Moving from Model 2 to Model 3, coefficients of log distance, squared log distance and the province neighbour see strong changes (Table 6.5). Controlling the East as the origin region and holding all the other covariates at their means, this sub-section plots Figure 6.4 to describe the relationship between log distance and log migration moving from Model 2 to 3. Note that fitted lines of the other three origin regions will be parallel to those of the East origin for neighbouring and non-neighbouring flows respectively as evidenced by Figure 6.4. Therefore, there is no need to further include figures of the other three origin regions to compare the marginal effect of log distance here. In Figure 6.4, to be specific, fitted lines of both neighbouring and non-neighbouring flows in Model 2 cut through those of Model 3 respectively from the bottom to the top, indicating that an increase in distance is associated

[104] Recalling the coefficient interpretation in Section 6.4 of Chapter 6, as the dependent variable is the natural log of the number of migrants and the neighbouring-province covariate is not on the log scale.

with a greater decrease in migration, when all the other covariates are held fixed, moving from Model 2 to 3. In other words, the marginal effect of distance decay upon migration increases moving from Model 2 to Model 3. Meanwhile, the coefficient of the province neighbour decreases from 0.623 in Model 2 to 0.241 in Model 3 (the unique effect when Region = 0, meaning the East) and becomes insignificant. All these changes show that Model 2 mistakenly underestimates effects of distance decay and the neighbouring province, by neglecting the effect of the provincial region along with its interactions with the neighbouring province.

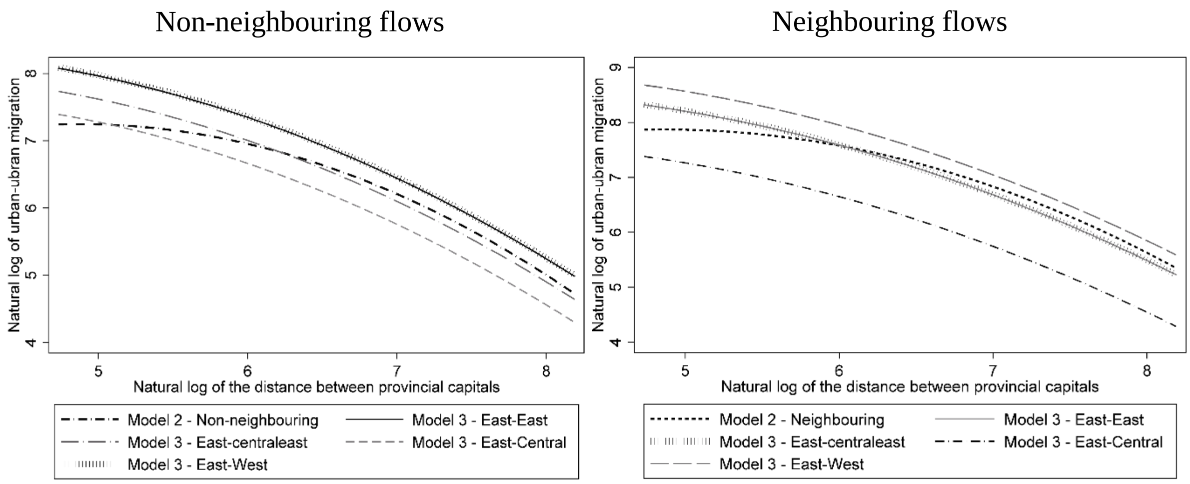

Figure 6.4 Comparison of fitted lines between Model 2 and 3 controlling the East as the origin region

Moreover, the constant, and coefficients of urban population and income in the destination, see further changes among all the common covariates moving from Model 2 to Model 3 (Table 6.5). Specifically, a 10% increase in the destination urban population is associated with a 6% increase in the volume of migration flows in Model 2 ($100 * (1.10^{0.6} - 1)\%$ = 6%), but an 8% increase in the migration flow volume in Model 3 ($100 * (1.10^{0.8} - 1)\%$ = 8%), all else equal. This illustrates that Model 2 underestimates the positive association between in-migration flows and destination urban population, which is evidently stronger. By contrast, Model 2 overestimates the positive association between destination urban income and in-migration flows. For instance, a 10% increase in destination urban income is associated with 22% growth in the volume of migration flows in Model 2 ($100 * (1.10^{2.2} - 1)\%$ = 22%), but is only associated with a 14% such volume increase in Model 3 ($100 * (1.10^{1.4} - 1)\%$ = 14%), all else equal. The observed changes of these coefficients signpost that introducing the provincial region and its relevant interactions has clearly influenced the relationships found in Model 2.

Regarding the unique effect of provincial regions on migration in Model 3, however, only the Central region is significantly different from the East in attracting migrants (coefficient (p value): -0.946 ($p < 0.01$)), although all three regions receives less in-flows than the East (Table 6.5). This can be observed in Figure 6.4: the fitted line of the East-Central region departs the most from that of the East-East for both neighbouring and non-neighbouring flows. Moreover, this effect also holds true for the other three origin regions when comparing the East and Central as the destination region respectively, recalling the parallel relationship of fitted lines for different origin regions (Figure 6.5)[105]. Specifically, the Central region receives 39% of the volume of in-flows of the East on average controlling all the other covariates, noting that the dependent variable is on the natural log scale and that the exponential of -0.946 is equal to 0.388 (100 * (1-exp(-0.946))% = 39%). This illustrates that the Central region is an evidently unpopular destination controlling all the other covariates, compared with the East. In terms of the origin region effect, the central and west regions export fewer migrants than the East controlling all the other covariates, but the differences are not statistically significant. By contrast, the Northeast sends out slightly more migrants than the East controlling all the other covariates (about 0.7%, as 100 * (exp(0.007)-1)% = 0.7%), whilst this difference is not significant either. Most importantly, the overall

[105] See Figure 6.7-6.9 in the Appendix.

model fit of Model 3 is significantly better than that of Model 2, as shown in former Wald tests[106]. Thus, even Model 2 proves insufficient in capturing the provincial region effects.

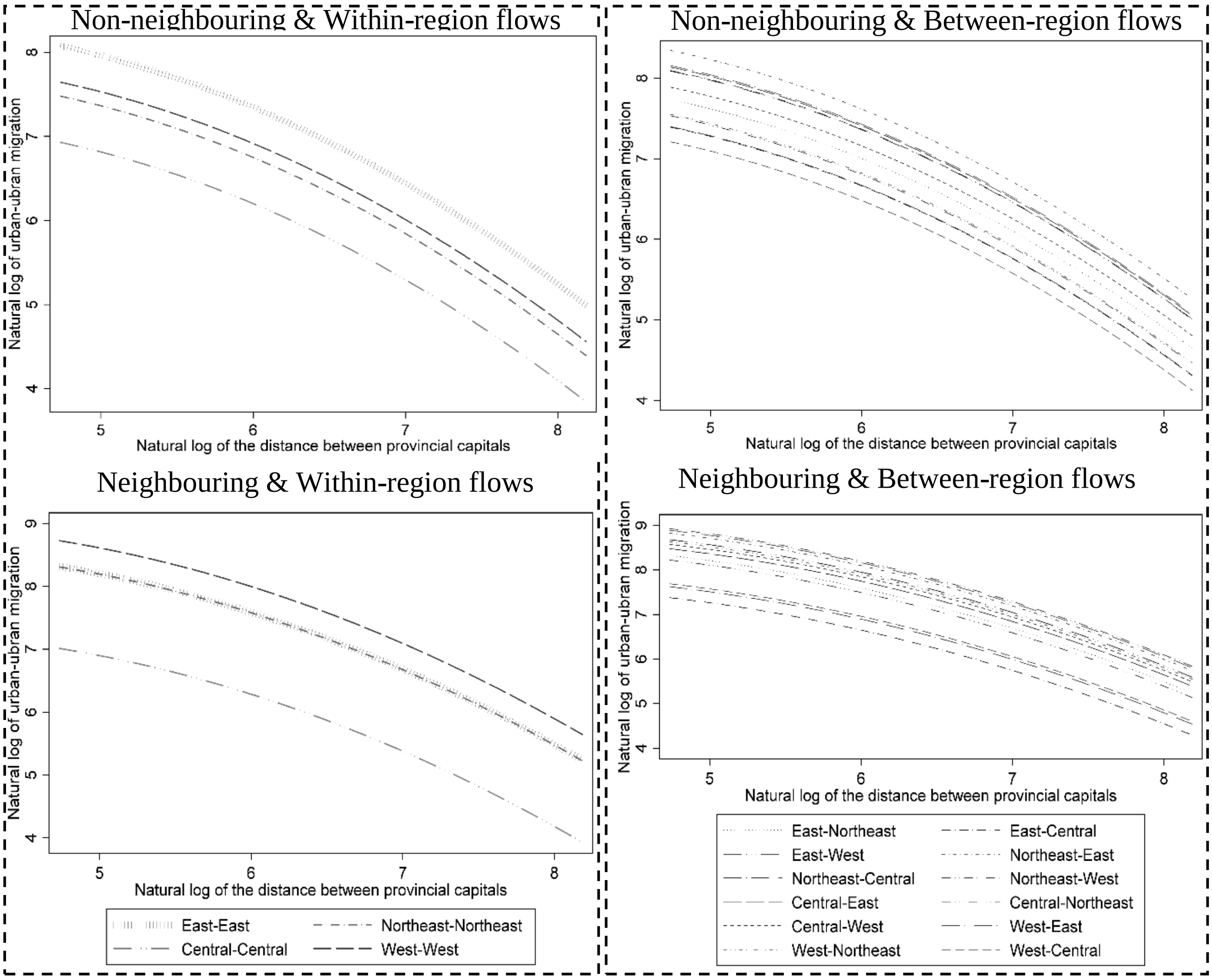

Figure 6.5 Comparison of fitted lines in Model 3 controlling the neighbouring-province and within-region covariates

As with the within-region effect, it is significantly negative for the East (Table 6.5). This negative effect also holds true for other regions. As observed in Figure 6.5, within-region flows are smaller than

[106] Recalling the significant Wald tests (the within-region covariate ($\chi_1^2 = 19.14, p < 0.001$), the destination region ($\chi_3^2 = 15.02, p < 0.01$), and the interaction between neighbouring province and origin region ($\chi_3^2 = 16.17$, $p < 0.01$) and that of neighbouring province and destination region ($\chi_3^2 = 23.28, p < 0.001$) in Model 3).

between-region ones for both neighbouring and non-neighbouring provinces respectively, holding all the other covariates at their means[107]. For instance, on average the volume of within-region flows is 77% of that of between-region flows (100 * (1-exp(-0.262))% = 77%) for the East, all else equal. This shows that migrants significantly prefer to migrate out of their origin regions regardless of the potentially high travel cost associated with the longer migratory distance, controlling all the other covariates. One possible explanation for this lies in that migrants are highly economically driven: provinces of the same region tend to be similar in socioeconomic conditions, so they need to travel beyond the region border to pursue farther and better economic opportunities.

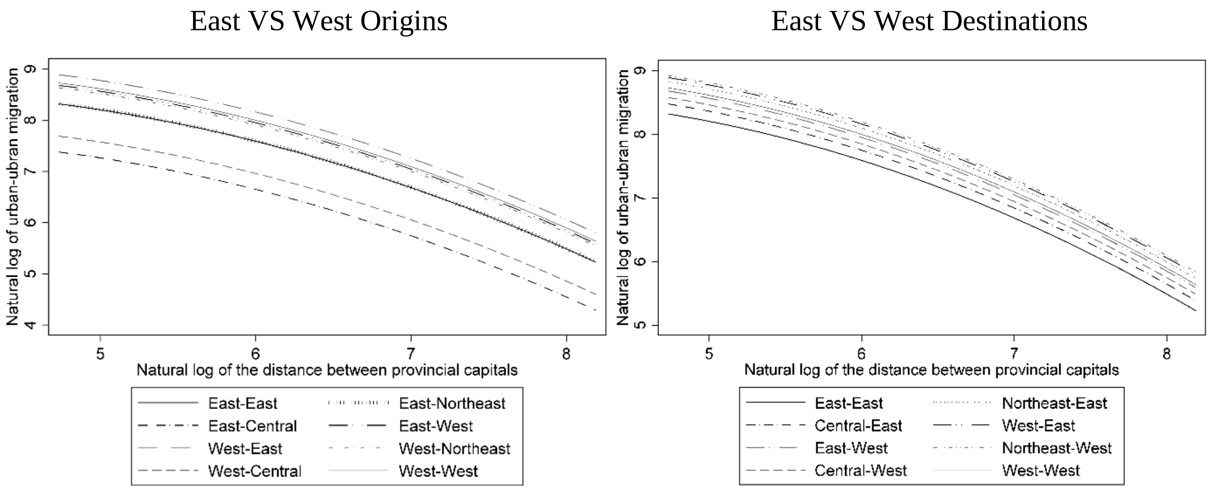

Figure 6.6 Comparison of fitted lines in Model 3 for neighbouring flows controlling the East and West as origins and destinations

Among all the six interaction terms, interactions of the neighbouring province and both the origin and destination West are significant (Table 6.5). Both interaction terms show that there are significant effects of the neighbouring province, whether the West is the origin or destination (Figure 6.6). To be more specific, out-flows of neighbouring provinces are 63% larger on average when the origin region is the West compared with those of the East (100*(exp(0.491)-1)%= 63%), controlling all the other

[107] More details can be seen from Figure 6.7-6.9 in the Appendix.

covariates (Figure 6.6). This indicates that the West has stronger exporting capabilities than the East for flows of neighbouring provinces originating from that region. By contrast, in-flows of neighbouring provinces are 41% larger on average when the destination region is the West rather than the East (100*(exp(0.349)-1)%= 41%), all else equal (Figure 6.6). This means that the West has stronger attraction for in-flows of neighbouring provinces.

Further Wald tests about the pairing covariates of origin and destination provinces reveal how the push and pull forces play out for urban-urban migration flows in Model 3. On the one hand, for instance, the effect sizes of the urban population of the origin and the destination becomes statistically indistinguishable (Wald test χ_1^2= 1.09, P = 0.296) in Model 3. This illustrates that the higher the destination population the greater the in-flows, and that the lower the origin population the smaller the out-flows. Indeed, it is the combined forces of the origin and the destination population that jointly facilitate the urban population movement between provinces. On the other hand, the effect sizes of the origin and the destination urban income remain significantly different and in opposite signs (χ_1^2= 15.47, P < 0.001). To be more specific, the destination urban income exerts strong attraction towards the urban-urban migration flows, whilst the origin urban income has a substantial negative association with out-flows.

In terms of the random parts, all three models from Model 1 to 3 have significant but distinctive destination province variances, whilst origin province and individual flow variances as well as flow-pair correlations remain relatively stable across models (Table 6.5). Specifically, the destination province variance drops from 0.24 in Model 1 to 0.22 in Model 2, which is brought about by the neighbouring-province covariate (Table 6.5). Then, it further decreases to 0.14 in Model 3 (Table 6.5), evidencing the notable change caused by the destination region dummy covariates. By contrast, the origin variance is relatively small, which is significant and decreases to different extents moving from Model 1 to 3 but less evidently. This means the origin effect is less sensitive to the contiguity of provinces and regions, compared with the destination effect. The individual flow variance, however, drops from 0.34 in Model 1 to 0.31 in Model 2, and then further decreases to 0.29 in Model 3 (Table 6.5). Meanwhile, the origin-destination correlation remains insignificant across models, although it also declines overall. This illustrates that both Model 1 and 2 overestimate the variances (particularly the destination province variance) by different degrees, when not considering effects of either neighbouring province or provincial region or both.

Additionally, the Appendix offers more detailed explanation about the random effects and how they change across models and provinces. Specifically, all three models of Model 1, 2 and 3 are similar in both their values and rankings of the provinces of Model 0 regarding the origin and the destination province effect respectively. Therefore, a brief summary of the observed variations across models is presented in the following, whilst more detailed interpretation of random effects on the provincial level can be found in the Appendix.

The origin province effect remains relatively stable in both the values and VPCs moving from Model 1 to 3, whilst by contrast the destination province effect is more variable in both the values and VPCs across models[108]. There are a few provinces significantly either above or below their regional averages in exporting and attracting migrants in Model 3. Among them, of particular interest is Heilongjiang, which is in the Northeast. It changes from exporting more urban migrants than the theoretical national average to not significantly different from the regional average, after controlling all the neighbouring-province and provincial-region covariates.

Moreover, the VPC of the individual flow variance increases moving from Model 1 to 3, by standing at 54% along with the significant correlation of flow-pairs (0.68, $p<0.001$) in Model 3. This is related to the reduction in the destination variance. As with the four types of the flow dependency, they all experience different degrees of decrease overall moving across all the models, except for the Type 1

[108] See Table 6.7 in Appendix.

dependency of common-origin flows. These changes in the four types of flow dependency are in line with those of the random effects.

6.6 Discussions

This section will focus on discussing the fixed-part effects first, and then move on to investigate the random-part effects. By doing so, it addresses the research questions proposed in the beginning of this chapter. Specifically, based on the results, this section will explain how migration flows between neighbouring and non-neighbouring provinces are different from each other, and in what ways migration flows of different regions are distinctive. The fixed effects reveal how different predictors affect the urban-urban migration, whilst random effects contribute to the understanding of how provinces systematically behave differently in exporting and attracting migrants. Overall, bringing the discussion of fixed and random effects together is vital to reveal how distance decay plays out in the regional migration system.

6.6.1 Fixed-part effects

The results from Model 3 echo previous findings to some extent, particularly the effect of populations and incomes of the origin and the destination as well as the distance. However, the major difference is that former studies have focused on total rather than a specific type of migration flows (such as urban-urban migration in this chapter), thus overlooking the neighbouring-province and provincial-region effects.

Regarding the effect of populations, previous studies on total migration in China have shown that the origin population is a more influential indicator for migration flows than the destination population (Fan 2005b), which also holds true for urban-urban migration. Nevertheless, the effect size difference of the urban population of the origin and the destination becomes insignificant (Wald test χ_1^2 = 1.09, P = 0.296), after controlling the neighbouring-province and provincial-region effects along with their interaction effects in Model 3. Neglecting those effects is shown, therefore, to lead to some misunderstanding of the migration system, recalling the underestimation of the destination population effect in both Model 1 and 2.

Effect sizes of the origin and destination urban income evidently converge when measured by absolute values, but they are significantly distinguishable when measured by true values (χ_1^2 = 15.47, P < 0.001) due to their opposite signs, all else equal. This is consistent with the consensus in the push-pull theory that the pull force from the destination wins over and thus creates in-flows (Dorigo and Tobler 1983; Mohtadi 1990). Specifically, destination urban income acts as a strong attractor pulling in the urban-urban migration flows, whereas by contrast origin urban income has a substantial negative association with out-migrants. This illustrates that the higher the destination income the greater the in-flows, and that the lower the origin income the greater the out-flows. Indeed, it is the income gap between the origin and the destination that underpins the population movement in pursuing economic returns (Crane 1992; Zhu 2002; Wang 2004; Schiff 2008).

The marginal effect of distance decay drops for distances below 1,765 km but increases above it for urban-urban migration flows, after further introducing the regional covariates on top of the neighbouring-province covariate. In practical terms, the turning point of 1,765 km means that it is the balancing point where the growth of travelling cost is equal to the increase in the likelihood of migration return (De Jong and Fawcett 1981; Evans 1990; De Haas 2010a; Meeus 2012; Molho 2013; Otoiu 2014). This supports the earlier theoretical suggestion of relaxing the linear relationship between log distance and log migration. Former studies have not fully explored the non-linearity of distance decay for urban-urban migration flows, although existing studies have found the negative association between distance and the total migration (Yan 2007; Shen 2013). Indeed, one tends to find a constant marginal effect of distance decay, if overlooking the potential non-linearity of distance decay.

Importantly, the average distance between provincial capitals is 1,380 km, and about 74% of all the 930 flows have below 1,765 km distances. This means that the majority of the flows undergo some decline in the marginal effect of distance decay, when fully considering both the neighbouring-province and provincial-region effects along with their interactions. This overall pattern of distance decay is likely to persist, though the actual value of turning point of the distance might fluctuate depending on different calibration methods. Recalling that the distance between provincial capitals is used in the models as a representative measurement of the actual migratory distance. This is because the census data is not presented at the individual level so the actual migratory distance is not available in the dataset, whilst the distance between provincial capitals has been proven to be an effective proxy measurement for the actual between-province migratory distance (Fan 2005b; Liu, Qi et al. 2015). Therefore, assuming a constant marginal effect of distance decay goes against the urban-urban interprovincial migration reality in China. Although China has made huge investments in domestic infrastructure and transportation to reduce the travelling cost during the past few decades (Luo, Zhu et al. 2014), this effect is not universal to all the flows. In fact, some long-distance migration (above 1,765 km) sees rises in the marginal effect of distance decay. To put it another way, it is advisable to invest more resources to improve the transport connection between faraway provinces, as this will result in more deduction on distance decay and lead to better between-province connectivity.

The unique provincial neighbourhood effect is positive but insignificant when controlling the region to be the reference region of the East. In other words, neighbouring provinces of the East may enjoy higher flows between themselves than those of the non-neighbouring provinces, all else being equal. This would echo the general finding in migration studies that people tend to be mobile in developed regions (De Haas 2010b; Molho 2013; Otoiu 2014). More importantly, considering the neighbouring relationship has improved the modelling fit for this study although this effect alone is not significant. In fact, Shen (2016a, 2016b) and Chun (2008) also demonstrated that considering the neighbouring relationship could greatly improve the model precision in predicting the total migration within China and the U.S. respectively.

Compared with those of the East, flows between neighbouring provinces are 63% and 41% larger on average when the West is the origin and the destination respectively, all else equal. However, migration flows are not significantly different between the West and East, if not considering the interaction effect with the neighbouring-province covariate. One possible explanation is that the neighbouring relationship represents both nearness and connectivity between provinces (Olsson 1965; LeSage 2008), which matters more for provinces of the West rather than the East. This reveals that the West has higher dependence upon the provincial nearness and connectivity, indicating provinces of the West are subject to more distance friction and need to rely more on neighbouring provinces to attract or export migration flows. This may be related to relatively larger sizes of the West region as well as its provinces (Figure 6.1). Moreover, the infrastructure and transportation system of the West is not as advanced as the East (Luo, Zhu et al. 2014), which may have caused this regional difference.

When comparing regions by investigating non-neighbouring provinces, the Central region receives significantly fewer migrants than the East having adjusted for the other factors in the model, whilst all the regions do not differ significantly in exporting migrants. The unpopularity of the Central region has roots in the unbalanced regional development policy (Démurger, Sachs et al. 2002; Fan, Kanbur et al. 2009; Sun 2013), and is both the cause and consequence of the spatially clustered migration flows (Groenewold, Chen et al. 2010; Liu, Stillwell et al. 2014). The Northeast also sends out slightly more migrants than the East for non-neighbouring provinces, controlling all the other factors. This is in line with the existing findings about the Northeast increasingly becoming the source of out-migration (Lee 2007; Chovanec 2009; Fleisher, Li et al. 2010; He, Bayrak et al. 2017).

The within-region effect is significantly negative for the East, on average incurring a 23% volume decrease of between-region flows when all the other covariates are held equal. This evidence lends strong support to including the regional covariates in the modelling. Specifically, it illustrates the extent to which the between-region movement is preferred by urban-urban migrants as a necessary choice in pursuit of better economic opportunities. Indeed, urban-urban interprovincial migrants in China are likely to be as economically driven as other (including some international) migrants around the world (Olsson 1965; Molho 2013; Otoiu 2014; Thomas, Stillwell et al. 2015). Since places within the same region are more likely to share similar socioeconomic characteristics, it is not surprising that migrants tend to move to places outside of the region with better economic opportunities as predicted by classical migration theories (De Haas 2010b; Molho 2013; Otoiu 2014).

6.6.2 Random-part effects

Including covariates for the neighbouring-province alone reduce the total residual variance by 9%, whilst considering both neighbouring-province and provincial-region effects reduces the total residual variance by 17%. This shows that both effects substantially improve the model fit. Specifically, all the random effect parameters, except for the origin-destination correlation coefficient, are significant and contribute to explaining the total residual. Moreover, adding neighbouring-province and provincial-region effects also changes the structure of VPCs, as both destination province and individual flow variances see notable drops.

Regarding the origin province effect, it remains relatively stable in both the values and VPCs moving from Model 1 to 3. Specifically, there are four provinces significantly below their regional-averages, among which Guizhou, Guangxi and Yunnan are all from the West whilst Shanxi is in the Central region. Urbanisation rate and economic structure might have accounted for this. Compared with the national average urbanisation rate of 50% and the regional average of 42% in 2010, the three provinces of the West all had relatively lower urbanisation rates ranging from 34% to 40%, which might lead to lower volumes of urban out-migrants (Davin 1996; Kundu and Gupta 1996; Chovanec 2009). Although the urbanisation rate of Shanxi (48%) was higher than the regional average of the West (44%) in 2010, its economy was primarily resource-oriented and labour-intensive such as coal mining, which might have offered plenty of local employment opportunities and thus reduced the out-migration (Chang and Dong 2016; Li, Lei et al. 2017; Li, Stoeckl et al. 2017).

Heilongjiang of the Northeast exports more urban migrants than the theoretical national average predicted by populations, incomes, distances and neighbouring provinces. However, such difference disappeared after introducing the neighbouring-province and provincial-region covariates. Heilongjiang has suffered from economic stagnation along with Jilin and Liaoning since a wave of privatization failures and factory closures in the Northeast in the 1990s (Cao, Qian et al. 1999; Han and Pannell 1999; Jiang 2004), and its urbanisation rate (56%) was slightly below regional-average of the Northeast (57%) in 2010. It suggests that the overestimated origin province effect of Heilongjiang is due to overlooking the regional effects of the Northeast, which is known as the rustbelt of China and undergoing brain-drain (Lee 2007; Chovanec 2009). These findings also lend support to the four-region division criterion raised in the earlier sections: neighbouring effects matter (Lupton 2003; Jenks and Dempsey 2007), and each region displays unique characteristics (Démurger, Sachs et al. 2002; Fleisher, Li et al. 2010; He, Bayrak et al. 2017).

By contrast, the destination province effect is more variable than that of the origin in both the estimates and VPCs, when controlling the effects of the neighbouring-province and provincial-region covariates. Specifically, there are five provinces with significantly below-region averages, among which Shandong, Jiangsu and Tianjin are all from the East whilst Tibet and Inner Mongolia are in the West, after controlling all the other covariates. As with the three provinces of the East, the urbanisation rate of Shandong (50%) was among the lowest in the region (the regional average of the East is 64%) in 2010, which might have affected its attraction. By contrast, Jiangsu and Tianjin might have suffered from the nearness and connectivity of two powerful megacities - Shanghai and Beijing respectively. Indeed, a low urbanisation rate represents a relatively small portion of urban economy and market in the provincial total, which is not particularly attractive towards in-migrants (Davin 1996; Kundu and Gupta 1996; Potts 2016). Moreover, the distance of Beijing and Tianjin is only 114 km, whilst that of Shanghai and Jiangsu is 276 km. Both were among the shortest distances (the national average is 1,380 km). Therefore, as the nearest neighbours of Shanghai and Beijing, the attractions of Jiangsu and Tianjin were diverted and overshadowed respectively (Head and Ries 1996; Hendrischke and Feng 1999).

Among the six provinces with significant above regional-average exporting capabilities, Xinjiang, Shaanxi and Ningxia are from the West whilst Hainan, Beijing and Guangdong are in the East. As expected, Xinjiang (43%), Shaanxi (46%) and Ningxia (48%) all had the above regional-average urbanisation rates among provinces of the West in 2010 (the regional-average urbanisation rate of the West is 42%), which contributed to their above regional-average attractiveness. Similarly, Beijing (89%) and Guangdong (66%) both had above regional-average urbanisation rate (the regional average of the East is 64%) in 2010, whilst Hainan's above regional-average attractiveness had little to do with its urbanisation rate (50%) but was more relevant to its pleasant climate and relatively cheap real estates (Gu and Wall 2007; Warner 2011; Sun, Chi et al. 2013; Inoue, Stickley et al. 2016; Wang, Wu et al. 2017).

Individual flow level VPC increased substantively after controlling the effects of the neighbouring-province and provincial-region covariates, which was primarily brought about by the reduction in the destination variance. Specifically, it accounted for 54% of the total residual, indicating that the majority of the variation in migration flows unexplained by the covariates relates to the unique interactions and relationships between pairs of provinces. This evidence is further supported by the correlation of flow-pairs, which stands at 0.68 and remains significant after controlling the effects of the neighbouring-province and provincial-region covariates. Indeed, the pairwise relationship is a fundamental element of urban-urban migration, which is consistent with other studies of pairwise data (Knight and Humphrey ; Gonzalez and Griffin 1997; Chun 2008). Nevertheless, this strong reciprocity between migration flow-pairs found at Chapter 5 may be an overestimate due to overlooking the provincial and regional neighbourhood effects of flows.

All types of the dependency experienced different degrees of decrease overall by controlling the effects of the neighbouring-province and provincial-region covariates, except for the Type 1 dependency of common-origin flows across all the models. Specifically, the Type 2 dependency of the common destination saw substantial decreases moving from Model 1 to 3, indicating that the overestimation of the Type 1 dependency type can be at least partly attributed to the neighbouring-province and provincial-region effects. In contrast, the Type 1 dependency of common-origin flows became stronger overall, suggesting that its underestimation is the result of neglecting the provincial-region effect. Both findings are consistent with the consensus in the scientific community that nearness and connectivity are critical components of socio-economic activities, which should be considered wherever possible (Lupton 2003; Jenks and Dempsey 2007). Then again, these patterns of change for the flow dependency corresponded with those of the random effects as they both declined, further confirming former interpretation about the random effects.

6.7 Conclusions

Using the multilevel gravity model of migration, this chapter has studied the nonlinearity of distance decay regarding urban-urban migration in China, by relaxing the assumption of linearity between log distance and log migration and introducing the neighbouring-province and provincial-region covariates.

The advanced modelling techniques proposed in this study have demonstrated that it is important to explore and examine the nonlinearity of distance decay for urban-urban migration. Indeed, the sophisticated model results offer an improved alternative to describe and investigate the complicated realities of the between-province urban population movement in China. Whilst previous research has provided great insight into China’s interprovincial migration, prior studies have tended to treat between-province migration flows as independent of each other and have struggled to fully investigate the four different flow independencies in the regional migration system. Neither does former research fully recognise nor consider the non-linearity of distance decay. Whilst the former chapter addressed the first knowledge gap, this chapter addresses the second.

Finally, this chapter provides a reformed theoretical framework for the investigation of the regional migration system. Indeed, flows are neither independent nor entirely obey the gravitation law of distance decay in the regional migration system, and further insights are needed to examine all four types of migration flows with the full dataset. Future research should endeavour to further the understanding of the relationship of migration and development for all four types of interprovincial migration flows in China.

Chapter 7: Conclusion

This chapter starts by highlighting the main findings of each results chapter. This final chapter therefore demonstrates how the former chapters work together to contribute to answering the overarching research question of the thesis, namely ‘what are the associations between China’s interprovincial migration and three levels of regional inequality (rural and urban, province and region)?’. It then details the main strengths and implications of this thesis by explaining potential applications in theoretical and empirical research and in practice such as policy making, as this thesis primarily contributes to methodological innovations and its application. This chapter subsequently argues for the contributions that the thesis has made to the research area. This chapter then discusses limitations of the data and methods utilised in the thesis. Finally, this chapter ends by pointing out directions of future research that deserve further investigations.

7.1 Main findings

This section summarises the main research findings from the three results chapters (Chapter 4, 5 and 6), by addressing each of the research questions in order:

(1) What are the associations between China’s interprovincial migration and rural and urban level of regional inequality through origin and destination population, income and distance?

(2) What are the associations between China’s interprovincial migration and province level of regional inequality through flow dependencies?

(3) What are the associations between China’s interprovincial migration and region level of regional inequality through distance decay?

Although each of these three chapters is a distinct piece of research with its own rationale and research question, together they have answered the overarching research question of the thesis with their own specific conclusions.

(1) The associations between China's interprovincial migration and rural and urban level of regional inequality are multi-directional, leading to four types of flows (rural-to-urban, urban-to-urban, rural-to-rural and urban-to-rural flows). These four flow types may differ from each other in terms of their causes.

Chapter 4 has answered the first research question by developing a new way of using the gravity model to separate out between-province migration flows into their rural and urban components. The results from Chapter 4 show that these four types of flow may differ from each other in terms of their causes, as the four different types of migration flows tend to have significantly different estimates for the same predictors and the four equations are strongly distinguishable from each other as a whole. Nevertheless, the results also show that destination urban income is always significantly positively associated with all four flows. By contrast, destination rural income remains significantly negatively associated, possibly due to the general undesirability of rural employment opportunities which are usually offered by low-paid and labour-intensive industries. Urban-rural and rural-urban flows have similar effect sizes for the same migratory distance, implying that both flows have to overcome similar institutional barriers in the rural and urban segregated Hukou system. This is in line with the general observation in real world context: urban and rural migrants are likely to encounter the same extent of institutional impediment when making between rural and urban movements, despite urban migrants' generally higher human capitals than their rural counterparts. Here, distance not only represents geographical friction but is also an indirect measurement of institutional or policy impacts (namely the Hukou system), answering an earlier speculation about the complicated role of distance in Chapter 3[109].

(2) Four flow dependencies (origin, destination, origin-destination, and flow-pair) affects interpretation of the associations between interprovincial migration and province level of regional inequality.

Chapter 5 proposes a multilevel gravity model to answer the second research question through examining interprovincial urban-urban migration. The flow dependencies derive from the interconnections between migration flows: between those that share an origin, those that share a destination, and where there is a reciprocal flow between places. The origin effect, the destination effect and the flow-pair effect are tested to be significant, with greater variations observed in the numbers of migrants received by provinces than in the numbers sent. This means that large numbers of urban migrants may tend to settle in popular destinations such as Beijing and Guangdong, whilst provinces tend to be less distinctive in their emissivity of urban migrants. Most importantly, reciprocal migration between pairs of provinces is an important feature of urban-urban migration. Similar to rural-urban migration, urban-urban migration is also strongly economically driven. Moreover, development level is likely to be closely linked with urban-urban migration. The degree of urbanisation plays a key role in provinces' origin effect whilst provinces of the least developed Western region tend to have the destination and pair-flow effects departing most from the national mean. As distance plays a sizeable deterring role, it has important policy implications for reducing regional inequality by encouraging migrants between cities in the western and inland areas.

(3) Region level of regional inequality affects distance decay of China's interprovincial migration. Distance does not have an entirely linear relationship with the interprovincial urban-

[109] See Sub-section 3.1.3 in Chapter 3 for more details.

urban migration flows, because neighbouring effects matter and the uniqueness of each region is not negligible.

Chapter 6 answers the final research question by exploring the non-linearity of distance decay in interprovincial urban-urban migration flows. This new approach has revealed the spatial distribution of migration flows in a more realistic way. First and foremost, distance does not have an entirely linear relationship with the interprovincial urban-urban migration flows, as the majority of the flows undergo decline in the marginal effect of distance decay. Therefore, assuming a constant marginal effect of distance decay goes against the urban-urban interprovincial migration reality in China. Secondly, overlooking the neighbouring-province and provincial-region effects could result in the inaccurate interpretation of results – particularly the underestimation of destination population effect. Considering the contiguity of both provinces and regions simultaneously could substantially improve the modelling fit instead. The East not only enjoys higher between-neighbour flows, but also has significantly larger volumes of between- than within- region flows. By contrast, the West has larger volumes of both in- and out-flows than the East for flows between neighbouring provinces, whilst the Central region receives significantly fewer migrants than the East for flows between non-neighbouring provinces. Overall, apart from the unpopularity of the Central region as a destination, provinces of the West are subject to more distance decay and need to rely more on neighbouring provinces to attract or export migration flows.

In answering these three research questions, step-by-step this thesis answers how China's interprovincial migration and regional inequality are associated at the levels of rural and urban, province and region. The results validate the development of novel data measurements and methods to cope with such issues in this thesis.

7.2 Research implications

Last section has summarised and discussed main findings of this thesis. Such findings have important research and policy implications. This section will discuss these implications for theoretical and empirical research and policies.

7.2.1 Theoretical and empirical implications

On the theoretical side, the thesis has shown clear resistance towards the possibility of a comprehensive analytical framework of migration that could explain all four flow types completely or equally well. This may seem contradictive to the general trend of converging analytical framework[110] occurring in migration theoretical perspectives at first sight. Underpinning this theoretical contrast, however, lies China's messy interprovincial reality, which departs fundamentally from a general theory that has to make too many assumptions to override countless inconvenient real-world variations. Equally important, this thesis does not claim that all four migration types bear too little linkage or resemblance to each other to be jointly theorised. That would be disintegrating a factually integral migration system into completely un-synthesised or unrelated components in a very narrow empiricist perspective. In that sense, this thesis actually has served as a middle path to advance migration theory between over-generalisation and pure-structuralism: seeing all four types as different but related leads towards a greater understanding of the nature and complexity of interprovincial migration in China.

On the empirical side, this thesis has highlighted the importance of flow dependencies. Such data dependencies are of both methodological and substantial interests, which has been explicitly explained by the proposed multilevel gravity model in the thesis. This multilevel gravity model of migration proposed in this thesis could also be used in other research fields such as trade and traffic flows (Bergstrand 1985; Guy 1987; Taaffe 1996). This is because trade and traffic flows share similar important attributes such as being bilateral and pairwise. By applying the proposed model to other

[110] See Section 2.1 in Chapter 2 for more details.

bilateral flow data, it will be of particular interest to reduce estimation bias by measuring and accounting for flow dependencies. In this sense, this thesis has contributed to the academic literature by demonstrating the importance of considering and accounting for unpicked data dependence in bilateral flow research.

7.2.2 Policy implications

Regarding internal affairs, the overarching goal of the state is to promote socio-economic growth and reduce regional inequality. To that end, the state has implemented a wide range of policies as mentioned earlier. Because China is in the process of fast industrialisation and urbanisation, managing labour migration has also been an essential part of internal affair governance. For that matter, this thesis offers three important policy implications as detailed below.

First and foremost, internal migration flows should not be treated as homogeneous by policy makers. Although the majority of migrants were city-oriented, there were already early signs of counter-urbanisation with the presence of the one million urban-rural migrants in 2010. This calls for further improving current migration policies, which should be tailored for different groups of internal migrants. The current migration policies primarily target at rural-urban migrants, whilst there is a general neglect of the needs of the other three migrant groups. The neglect towards urban-rural migrants, however, is particularly damaging in time of a fast-aging China, as policy makers and society are relatively un-warned and unprepared about the magnitude and impact of urban-rural migrants. Results of this thesis have shown they are less likely economically driven, although it is still unclear about their other attributes. Research in other countries has revealed that the majority of urban-rural migrants are retirees. It is conceivable that Chinese urban-rural migrants are likely to be retirees as well. Coupled with the fast aging population in China, developing such policies therefore seems of urgent need. Urban-rural migration research in developed countries has highlighted their amenity-seeking behaviour with nostalgia for the countryside. It is therefore recommended to policy makers of China to look into developing proper supportive system in rural areas, such as retirement supportive schemes, hospitals and nursing homes.

Secondly, the current migration management and regional development policies may have been mismatching with each other for a few remote provinces. In implementing China Western Development (a major regional development policy[111]), the state has initiated multiple large-scale infrastructure development projects in remote provinces such as Xinjiang, Yunnan and Tibet, in the hope of improving their connectivity with the socio-economic heartland (i.e. the eastern coastal region) to facilitate socio-economic growth. With the improvement of infrastructure and connectivity, however, these remote provinces have seen inevitably rising out-migration and brain-drain, despite the local favourable population policies such as increasing wages for skilled workers. This contradicts with the general goal of migration management policies of the state, which has been encouraging immigration to remote provinces for both regional equality and national security reasons. This is exemplified by remote provinces becoming unpopular destinations but increasingly popular origins from 2000 to 2010 for total migration overall. To resolve this policy mismatch, the state could implement stronger in-flow encouraging policies such as provision of housing subsidies to mitigate the rising out-migration. In doing so, the state also needs to be aware that these provinces tend to have vulnerable environments and fragile ecosystems. It is therefore advisable to set a reasonable population limit for each of such provinces in accordance with their human carrying capacities.

The third policy implication relates to the nationwide goal of localised urbanisation, which has been controlling the growth and sprawl of mega cities and encouraging immigration to medium and small cities. Such policy goal was at least partly failed between 2000 and 2010, exemplified by the growing popularity of mega cities such as Beijing and Shanghai during this period along with their particular attractiveness as urban destinations for both urban- and rural-origin migrants in 2010. Such agglomeration of population and industries has led to environment and governance difficulties such as serious air pollution and severe traffic congestions in mega cities. Many such cities have therefore

[111] See Sub-section 2.3.1.3 in Chapter 2 for more details.

strengthened their local Hukou systems in the hope for reducing in-flows, which has been largely unsuccessful to this day. On account of this, combined measurements may be beneficial. For instance, the state could engage in a wide range of policies including stricter and proper urban planning of mega cities, moving labour-intensive manufacturing industries to suburban areas or nearby medium and small cities, and encouraging rural migrants to return to origins to start their own businesses by providing capital and market support from destination cities.

In summary, China's internal flows are spatially unbalanced and featured by migrant groups differing in migratory behaviours and experiences, whilst migrant management policies should always be made to meet needs of migrants and compatible with economic development. It is vital that China must learn from other countries' experience in redistributing population with a much greater degree of movement freedom, with the ultimate aim to remove institutional or social barriers that hinder people's right of pursuing a better life. To advise on future governance of China, research on internal migration in other countries offers valuable methods and theories that could be used in China's case. In that sense, state support for future research upon internal migration holds great value and is of constant need.

7.3 Limitations

Having discussed specific implications of this research, it must be acknowledged that this thesis is not without limitations. Whilst specific limitations have been explained within each results chapter, this section now discusses the main methodological limitations of this thesis as a whole, highlighting how these may impact the conclusions overall.

First, the set of variables used in the thesis is relatively limited. This is because the methodological aim of this thesis is to develop new extensions of the gravity model, so the selection of predictors has stuck to principles of traditional gravity models in order to present extended models in an as simple and accessible form as possible. Whilst key elements of the gravity model (population, income and distance) were included and investigated in models of this thesis, not all models of four flow types showed the same good extent of model fit. Model for urban-rural migration, for instance, was particularly less responsive to the selected predictors, revealing that urban-rural migrants were not entirely driven by income factors. Therefore, there is clear potential in expanding the set of predictors to further improve model fit.

Using the Hukou status as the defining factor of different migration types is the second limitation of this thesis. This limitation is two-fold. On the one hand, it leads to the necessary interpretation of migration data as (lifetime) migration flow rather than (yearly) cross-sectional migration flow, as the census data is effectively unable to count multiple and return moves[112]. To illustrate, urban-urban migration flow from Beijing to Shanghai in 2010 represents the stock of all urban migrants originating from Beijing and settling in Shanghai by the year of 2010. It therefore would be problematic to interpret this number as the migrants from Beijing to Shanghai between (for instance) 2009 and 2010. This interpretation limitation of response covariates may have contradicted with the cross-sectional selection of data points for key predictors. In estimating the urban-urban migration flow from Beijing to Shanghai in 2010, for example, this thesis selects 2010's population and income predictors (i.e. the final-year data points of migration stock). While such selection standard remains as an established and effective practice in academic studies (Fan 2005b; Shen 2015), it is conceivable that this might have biased results in this thesis. On the other hand, using the Hukou status as the defining factor of different migration types results in the binary classification of space into rural and urban areas[113]. Whilst such conceptual simplification is necessary in extracting data and developing measurements, it contradicts with the continuum of rural and urban areas in reality. Therefore, it may have induced bias in the results. This is exemplified by the definition of 'Town' in the Census, which consists of both urban ('Neighbourhood committees of the town') and rural ('Village committees of the town') areas. Using the Hukou status to divide a town into its rural and urban segments is arguably arbitrary, as both segments are essentially located in the same location.

[112] The census data collection practice along with its impact upon migration counts is extensively explained in Sub-section 3.1.1.

[113] The rural-urban continuum is explicitly illustrated with Figure 3.2 in Sub-section 3.1.1.

The third limitation lies in the missing role of intraprovincial migration throughout model specifications of this thesis. Although the Census does have out- and in-migration totals for each province, it offers no clue about where migrants are from or where they go within the province. It is therefore impossible to calculate the within-province migratory distance. Solving this requires migration data of county-level, which is the lower administrative unit of province in China. However, origin and destination information of such data is unavailable either in the Census or any other alternative datasets. This data limitation may have induced bias in the results, as effects of intraprovincial migration are underrepresented or unmeasured throughout all models of this thesis. Whilst Chapter 4 established its models to examine associations of four migration types through inexplicitly considering influences of intraprovincial migration, Chapter 5 and 6 set their focus on urban-urban interprovincial migration alone and thus do not account for any effects of within-province migration as an indication of associations between different types of migration flows.

The fourth limitation relates to model assumptions throughout the thesis. The first important model assumption was the independence of the total residuals. Whilst this assumption was relaxed in models from Chapter 4 to Chapter 5 and 6, approximate normality is still needed for the predicted random effects. The second assumption related to spatial autocorrelation. Although flow dependencies and neighbouring-province relationships may have partly helped explained away some spatial autocorrelation of migration flows, models of this thesis did not explicitly consider such effects. Entering key predictors linearly in models was the third important assumption. Although this assumption was made to present the extensions of the gravity model in as simple an accessible form as possible, effects of key predictors (i.e., population and distance) may taper off as provinces get larger. This could contradict with this assumed linear relationships between predictors and response covariates. Whilst Chapter 6 has explored the non-linearity of distance decay, similar investigations need to be conducted on population and income predictors. Lastly, this thesis has chosen linear over Poisson formulation of models. The main reason lay in the large average volume of interprovincial migration flows in China. It therefore needs to acknowledge that the Poisson formulation of the migration model conveys significant advantages over the linear regression formulation in applications where migration flows are far smaller (e.g., when carrying out migration studies at a finer spatial scale).

Finally, this thesis is not able to examine all the four types of migration flows with the full dataset due to the limitation of time and resources. This is particularly related to computational capacities of statistical software packages such as MLwiN (its capacity is determined entirely by the memory of the PC) and Stata (the maximum matrix size is 11,000 in Stata SE), as they set further limits on the exploration of multivariate multilevel gravity models for all four migration types. For instance, conducting multivariate multilevel gravity models for 31 provinces across all four migration types requires a matrix size of 61,504 to execute runmlwin in Stata[114].

7.4 Future work

The former section has addressed the limitations of this thesis, pointing out both empirical and methodological directions of future research that deserve further investigations. By following this guiding framework, this section will explicitly discuss both future research directions in order, highlighting potentially fruitful areas in need of more scholarly endeavour to deepen understanding of internal migration in China.

On the empirical side, future research needs to expand the set of predictors and investigate all four types of interprovincial migration flows, and to examine associations of intraprovincial and interprovincial migration flows. Economic factors such as employment rate and share of primary sector may have the potential for improving model fit, so do predictors beyond the scope of the economic perspective such as air quality and greenspace availability (Liu and Shen 2014a, 2014b; Shen 2016a). Whilst expanding investigations to all four types of interprovincial flows is highly promising for future studies, adding 'town' and/or 'township' as an intermediate unit between rural and urban areas is particularly interesting

[114] The calculation is (((31 origins + 31 destinations) * (4 flow types)) ^2=61,504). This is because the multivariate model is formulated as three levels in each equation for the 4 flow types in MLwiN, with lower level of 930 individual flows and two higher parallel levels of 31 origins and 31 destinations. It is the cumbersome nature of this formulation that prohibits the calculation.

and may offer fruitful research outputs at a finer scale of the rural-urban continuum division to better describe the reality in China. Additionally, the impact of intraprovincial migration may have had on model results of this thesis can be assessed by studies conducted on other populations. Research on internal migration of developed countries such as the U.S. and the U.K. can be of particular interest with higher qualities of data. If such studies do not suffer from the missing data issue, then consistency in findings may provide evidence of the direction and magnitude of bias in this thesis that is attributable to not accounting for intraprovincial migration flows.

On the methodological side, future work on migration flows should also seek to employ a range of alternative modelling strategies to further relax model assumptions such as accounting for effects of spatial autocorrelation. One potentially fruitful area that future studies may be able to investigate is to combine established spatial analysis models such as spatial filtering and spatial lag models with the multilevel gravity model of migration proposed by this thesis (Chun 2008; Crowder and South 2008). To illustrate, spatial filtering with eigen vectors could efficiently remove effects of spatial autocorrelation (Chun 2008), while spatial lag models explicitly measure spatial interactions of substantive empirical importance (Crowder and South 2008). These spatial analytical methods deal with spatial autocorrelation in different ways and have different sets of underlying assumptions and limitations to the modelling approaches. These differences are important given that migration is an equilibrium outcome of complicated social and spatial interaction processes contextualised by socio-economic, political and geographical conditions of specific regions. Another future research area is to further explore non-linearity of other (population and income) predictors. One way to do this is to simply enter origin and destination population into the model as polynomial functions (e.g., a quadratic or cubic relationship) rather than as linear only terms. Additionally, it is equally important to keep pursuing alternative ways to compute the multivariate formulation of the multilevel gravity model proposed by this thesis. For instance, it is worth to explore specifying the model in Stan to avoid the cumbersome formulation used in MLwiN. Advances in new technologies may facilitate this endeavour. Lastly, microsimulation models might offer new insights into measuring impacts of intraprovincial flows upon interprovincial migration (Clarke 1986; Acharya and Leon-Gonzalez 2013). If results from differing microsimulation modelling strategies are consistent, then the simulated county-level intraprovincial flows can be used to examine interprovincial migration in a more robust way. To illustrate, this can be achieved by drawing lessons from the social relations model which would be able to handle intraprovincial and interprovincial flows simultaneously (Kenny and Kashy 2011; Koster and Leckie 2014). It would then be able to directly assess estimation bias of the proposed model, and confidence in the conclusions may become stronger.

Both empirical and methodological directions identified here have great potential in improving knowledge of migration flows. Migration is clearly a complex social phenomenon, being influenced and impacted by a myriad of socioeconomic and demographic factors. It therefore requires the careful use of detailed datasets and appropriate statistical methods in future studies.

References

Acharya, C. P. and R. Leon-Gonzalez (2013). "The impact of remittance on poverty and inequality: A micro-simulation study for Nepal." Asian Journal of Empirical Research **3**(9): 1061-1080.

Afridi, F., S. X. Li, et al. (2015). "Social identity and inequality: The impact of China's hukou system." Journal of Public Economics **123**: 17-29.

Agadjanian, V., E. Gorina, et al. (2014). "Economic Incorporation, Civil Inclusion, and Social Ties: Plans to Return Home Among Central Asian Migrant Women in Moscow, Russia." International Migration Review **48**(3): 577-603.

Aghion, P., P. Howitt, et al. (1998). Endogenous growth theory, MIT press.

Agnew, J. A., V. Mamadouh, et al. (2015). The Wiley Blackwell companion to political geography. Chichester, UK ; Hoboken, NJ, John Wiley & Sons.

Ai, C. and E. C. Norton (2003). "Interaction terms in logit and probit models." Economics letters **80**(1): 123-129.

Alcoff, L. (1988). "Cultural feminism versus post-structuralism: The identity crisis in feminist theory." Signs: Journal of women in culture and society **13**(3): 405-436.

Alm, J. and J. V. Winters (2009). "Distance and intrastate college student migration." Economics of Education Review **28**(6): 728-738.

Alonso, W. (1986). "Systemic and log-linear models: from here to there then to now and this to that."

Andersson, A., G. Haag, et al. (2012). Interregional migration: dynamic theory and comparative analysis, Springer Science & Business Media.

Anh, N. T., J. Rigg, et al. (2012). "Becoming and being urban in Hanoi: Rural-urban migration and relations in Viet Nam." Journal of Peasant Studies **39**(5): 1103-1131.

Anselin, L. (2002). "Under the hood issues in the specification and interpretation of spatial regression models." Agricultural economics **27**(3): 247-267.

Appadurai, A. (1990). "Disjuncture and difference in the global cultural economy." Theory, culture & society **7**(2-3): 295-310.

Arato, A. (1993). From neo-Marxism to Democratic theory: essays on the critical theory of Soviet-type societies, ME Sharpe.

Arya, S., D. M. Mount, et al. (1998). "An optimal algorithm for approximate nearest neighbor searching fixed dimensions." Journal of the ACM (JACM) **45**(6): 891-923.

Aybek, C. M. (2015). "Time Matters: Temporal Aspects of Transnational Intimate Relationships and Marriage Migration Processes From Turkey to Germany." Journal of Family Issues **36**(11): 1529-1549.

Bagnai, A. (2009). "The role of China in global external imbalances: Some further evidence." China Economic Review **20**(3): 508-526.

Bailey, N. and M. Livingston (2008). "Selective migration and neighbourhood deprivation: Evidence from 2001 census migration data for England and Scotland." Urban Studies **45**(4): 943-961.

Banasiak, J. and A. Goswami (2015). "Singularly Perturbed Population Models with Reducible Migration Matrix 1. Sova-Kurtz Theorem and the Convergence to the Aggregated Model." Discrete and Continuous Dynamical Systems **35**(2): 617-635.

Bao, S., A. Shi, et al. (2005). "Zhongguo renkou qianyi de kongjian xingtai bianhua fenxi." Chinese Journal of Population Science **2005**(5): 28-36.

Bao, S. M., O. B. Bodvarsson, et al. (2011). "The Regulation of Migration in a Transition Economy: China's Hukou System." Contemporary Economic Policy **29**(4): 564-579.

Bao, S. M., G. H. Chang, et al. (2002). "Geographic factors and China's regional development under market reforms, 1978-1998." China Economic Review **13**(1): 89-111.

Barro, R. T. and X. Sala-i-Martin (1990). Economic growth and convergence across the United States, National Bureau of Economic Research.

Barro, R. T. and X. Sala-i-Martin (1992a). "Convergence." Journal of political Economy: 223-251.

Barro, R. T. and X. Sala-i-Martin (1992b). "Regional growth and migration: A Japan-United States comparison." Journal of the japanese and International Economies **6**(4): 312-346.

Bartholomew, D. J., F. Steele, et al. (2008). Analysis of multivariate social science data, Chapman and Hall/CRC.

Batty, M. and S. Mackie (1972). "The calibration of gravity, entropy, and related models of spatial interaction." Environment and Planning A **4**(2): 205-233.

Bavaud, F. (1998). "Models for spatial weights: a systematic look." Geographical analysis **30**(2): 153-171.

Becker, G. S. (1965). "A Theory of the Allocation of Time." The economic journal: 493-517.

Becker, G. S. (1994). Human capital revisited. Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition), The university of Chicago press: 15-28.

Becker, J. (2003). "On the move at last (New set of control on China's internal migration)." Index on Censorship **32**(2): 120-124.

Becquelin, N. (2000). "Xinjiang in the Nineties." The China Journal(44): 65-90.

Beine, M., S. Bertoli, et al. (2014). "A practitioners' guide to gravity models of international migration." Documentos de trabajo (FEDEA)(3): 1-27.

Bell, M., M. Blake, et al. (2002). "Cross‐national comparison of internal migration: issues and measures." Journal of the Royal Statistical Society: Series A (Statistics in Society) **165**(3): 435-464.

Bell, M. and S. Muhidin (2009). "Cross-national comparison of internal migration."

Berger, D. (2009). "Taxes, institutions and local governance: evidence from a natural experiment in colonial Nigeria." Unpublished manuscript.

Bergstrand, J. H. (1985). "The gravity equation in international trade: some microeconomic foundations and empirical evidence." The review of economics and statistics: 474-481.

Bernstein, H. (2010). Class dynamics of agrarian change, Kumarian Press.

Berry, B. J. (1980). "Urbanization and counterurbanization in the United States." The Annals of the American Academy of Political and Social Science **451**(1): 13-20.

Berry, B. J. L. (1976). Urbanization and counter-urbanization, SAGE Publications, Incorporated.

Bhat, P. N. M. (2002). "General growth balance method: A reformulation for populations open to migration." Population Studies-a Journal of Demography **56**(1): 23-34.

Bhattacharya, P. C. (2002). "Urbanisation in developing countries." Economic and political Weekly: 4219-4228.

Biao, X. (2007). "How far are the left‐behind left behind? A preliminary study in rural China." Population, Space and Place **13**(3): 179-191.

Black, R., W. N. Adger, et al. (2011). "The effect of environmental change on human migration." Global environmental change **21**: S3-S11.

Blacklock, C., C. Heneghan, et al. (2012). "Effect of UK policy on medical migration: a time series analysis of physician registration data." Human Resources for Health **10**.

Boccagni, P. and L. Baldassar (2015). "Emotions on the move: Mapping the emergent field of emotion and migration." Emotion, Space and Society **16**: 73-80.

Boccaletti, S., V. Latora, et al. (2006). "Complex networks: Structure and dynamics." Physics reports **424**(4-5): 175-308.

Boguná, M., R. Pastor-Satorras, et al. (2004). "Models of social networks based on social distance attachment." Physical review E **70**(5): 056122.

Borjas, G. J. (1989). "Economic theory and international migration." International migration review **23**(3): 457-485.

Bosker, M., S. Brakman, et al. (2012). "Relaxing Hukou: Increased labor mobility and China's economic geography." Journal of Urban Economics **72**(2-3): 252-266.

Boswell, C. (2008). "Combining Economics and Sociology in Migration Theory." Journal of Ethnic and Migration Studies **34**(4): 549-566.

Bottero, W. (2004). "Class identities and the identity of class." Sociology **38**(5): 985-1003.

Breman, J. (1978). "Seasonal Migration and Cooperative Capitalism - Crushing of Cane and of Labor by the Sugar Factories of Bardoli, South Gujarat .1." Journal of Peasant Studies **6**(1): 41-70.

Breman, J. (1979). "Seasonal Migration and Cooperative Capitalism 2. Crushing of Cane and of Labor by the Sugar Factories of Bardoli, South-Gujarat." Journal of Peasant Studies **6**(2): 168-209.

Brettell, C. B. and J. F. Hollifield (2014). The Politics of International Migration How Can "We Bring the State Back In"? Migration theory, Routledge**:** 239-300.

Brown, L. A. and E. G. Moore (1970). "The intra-urban migration process: a perspective." Geografiska Annaler: Series B, Human Geography **52**(1): 1-13.

Brown, L. A., J. Odland, et al. (1970). "Migration, Functional Distance, and Urban Hierarchy." Economic Geography **46**(3): 472-485.

Brubaker, R. and F. Cooper (2000). "Beyond "identity"." Theory and society **29**(1): 1-47.

Brunarska, Z., J. Nestorowicz, et al. (2014). "Intra- vs. extra-regional migration in the post-Soviet space." Eurasian Geography and Economics **55**(2): 133-155.

Buhaug, H. and K. S. Gleditsch (2008). "Contagion or confusion? Why conflicts cluster in space." International Studies Quarterly **52**(2): 215-233.

Buu, A., M. Mansour, et al. (2007). "Alcoholism effects on social migration and neighborhood effects on alcoholism over the course of 12 years." Alcoholism-Clinical and Experimental Research **31**(9): 1545-1551.

Byrd, W. A., W. Byrd, et al. (1990). China's rural industry: Structure, development, and reform, Oxford University Press.

Cai, F. and D. Wang (2003a). "Migration as marketization: what can we learn from China's 2000 census data?" China Review **3**(2): 73-93.

Cai, F. and D. Wang (2003b). "Zuowei shichanghua de renkou liudong diwuci renkou pucha shuju fenxi." Zhongguo renkou kexue **2003**(5): 11-19.

Cameron, A. C., J. B. Gelbach, et al. (2011). "Robust Inference With Multiway Clustering." Journal of Business & Economic Statistics **29**(2): 238-249.

Campbell, L. and D. A. Kashy (2002). "Estimating actor, partner, and interaction effects for dyadic data using PROC MIXED and HLM: A user–friendly guide." Personal Relationships **9**(3): 327-342.

Cao, G. Z., K. Li, et al. (2017). "Consumption Structure of Migrant Worker Families in China." China & World Economy **25**(4): 1-21.

Cao, Y., Y. Qian, et al. (1999). "From federalism, Chinese style to privatization, Chinese style." Economics of Transition **7**(1): 103-131.

Carling, J. and F. Collins (2018). Aspiration, desire and drivers of migration, Taylor & Francis.

Castles, S. (2002). "Migration and community formation under conditions of globalization." International migration review **36**(4): 1143-1168.

Castles, S., H. De Haas, et al. (2013). The age of migration: International population movements in the modern world, Palgrave Macmillan.

Cerrutti, M. and R. Bertoncello (2003). "Urbanization and internal migration patterns in Latin America." Centro de Estudios de Población. Argentina.

Champion, A. (1999). "Urbanization and counter-urbanization." Applied Geography. Principle and Practice: 347-357.

Champion, A. G. (1989). "Counterurbanization: the changing pace and nature of population deconcentration."

Champion, A. G. and D. Atkins (1996). The counterurbanisation cascade: an analysis of the 1991 Census Special Migration Statistics for Great Britain, Department of Geography University of Newcastle upon Tyne.

Chan, J. and N. Pun (2010). "Suicide as protest for the new generation of Chinese migrant workers: Foxconn, global capital, and the state." The Asia-Pacific Journal **37**(2): 1-50.

Chan, K. W. (2007). "Misconceptions and complexities in the study of China's cities: Definitions, statistics, and implications." Eurasian Geography and Economics **48**(4): 383-412.

Chan, K. W. (2009). "The Chinese hukou system at 50." Eurasian geography and economics **50**(2): 197-221.

Chan, K. W. (2010a). "The Global Financial Crisis and Migrant Workers in China: 'There is No Future as a Labourer; Returning to the Village has No Meaning'." International Journal of Urban and Regional Research **34**(3): 659-677.

Chan, K. W. (2010b). "The Household Registration System and Migrant Labor in China: Notes on a Debate." Population and Development Review **36**(2): 357-364.

Chan, K. W. (2013). "China, internal migration." The encyclopedia of global human migration **2**: 980-995.

Chan, K. W., T. Liu, et al. (1999). "Hukou and non‐hukou migrations in China: comparisons and contrasts." International Journal of Population Geography **5**(6): 425-448.
Chan, K. W. and L. Zhang (1999). "The hukou system and rural-urban migration in China: Processes and changes." The China Quarterly **160**: 818-855.
Chan, S. H. and H. H. Qiu (2011). "Loneliness, job satisfaction, and organizational commitment of migrant workers: empirical evidence from China." International Journal of Human Resource Management **22**(5): 1109-1127.
Chang, G. (2002). "The cause and cure of China's widening income disparity."
Chang, H. Q., X. Y. Dong, et al. (2011). "Labor Migration and Time Use Patterns of the Left-behind Children and Elderly in Rural China." World Development **39**(12): 2199-2210.
Chang, Y. Z. and S. C. Dong (2016). "Evaluation of Sustainable Development of Resources-Based Cities in Shanxi Province Based on Unascertained Measure." Sustainability **8**(6).
Che, Y. (2014). "Mismatch Land reallocations, recovery land rental and land rental market development in rural China." China Agricultural Economic Review **6**(2): 229-247.
Chen, A. P. and N. Groenewold (2011). "Regional Equality and National Development in China: Is There a Trade-Off?" Growth and Change **42**(4): 628-669.
Chen, H. X. (2011). "Land Property Right System Reformation and Sustainable Development." Proceedings of 2011 International Conference on Construction and Real Estate Management, Vols 1 and 2: 334-337.
Chen, R. S., C. Ye, et al. (2014). "The impact of rural out-migration on land use transition in China: Past, present and trend." Land Use Policy **40**: 101-110.
Chen, S. and J. L. Powell (2012). Aging in China: Implications to social policy of a changing economic state, Springer Science & Business Media.
Chen, S. H. and M. Ravallion (2004). "How have the world's poorest fared since the early 1980s?" World Bank Research Observer **19**(2): 141-169.
Chen, X. M. (1991). "China City Hierarchy, Urban-Policy and Spatial Development in the 1980s." Urban Studies **28**(3): 341-367.
Chen, Z., X. F. Liu, et al. (2013). "Beyond Lewis: rural-to-urban migration with endogenous policy change." China Agricultural Economic Review **5**(2): 213-230.
Chen, Z. H., Y. Ge, et al. (2013). "Globalization and Gender Wage Inequality in China." World Development **44**: 256-266.
Cheng, C. (1991). "Internal Migration in Mainland China - the Impact of Government Policies." Issues & Studies **27**(8): 47-70.
Cheng, Z. M., I. Nielsen, et al. (2014). "Access to social insurance in urban China: A comparative study of rural-urban and urban-urban migrants in Beijing." Habitat International **41**: 243-252.
Chiang, Y. L., E. Hannum, et al. (2014). "It's Not Just About the Money: Gender and Youth Migration from Rural China." Chinese Sociological Review **47**(2): 177-201.
Chovanec, P. (2009). "The nine nations of China." The Atlantic.
Christian, W. and W. Braden (1966). "Rural Migration and Gravity Model." Rural Sociology **31**(1): 73-80.
Chun, Y. (2008). "Modeling network autocorrelation within migration flows by eigenvector spatial filtering." Journal of Geographical Systems **10**(4): 317-344.
Claeson, C. F. (1969). "2-Stage Model of in-Migration to Urban Centres - Deductive Development of a Variant of Gravity Formulation." Geografiska Annaler Series B-Human Geography **51**(2): 127-138.
Clark, W. A. (1982). "Recent research on migration and mobility: a review and interpretation." Progress in planning **18**: 1-56.
Clarke, M. (1986). "Demographic processes and household dynamics: a microsimulation approach."
Cobb, C. W. and P. H. Douglas (1928). A theory of production. Proceedings of the Fortieth Annual Meeting of the American Economic Association.
Cohen, J. E., M. Roig, et al. (2008). "International migration beyond gravity: A statistical model for use in population projections." Proceedings of the National Academy of Sciences of the United States of America **105**(40): 15269-15274.
Cong, Z. and M. Silverstein (2011). "Intergenerational Exchange Between Parents and Migrant and Nonmigrant Sons in Rural China." Journal of Marriage and Family **73**(1): 93-104.

Connell, J., B. Dasgupta, et al. (1976). "Migration from rural areas: the evidence from village studies."

Converse, P. D. (1949). "New laws of retail gravitation." Journal of marketing **14**(3): 379-384.

Conway, D. and J. H. Cohen (1998). "Consequences of migration and remittances for Mexican transnational communities." Economic geography **74**(1): 26-44.

Courgeau, D. and B. Baccaini (1989). "Migration and Distance." Population **44**(3): 659-663.

Crane, R. (1992). "Voluntary Income-Redistribution with Migration." Journal of Urban Economics **31**(1): 84-98.

Cristobal-Campoamor, A. (2009). "International convergence and local divergence." Annals of Regional Science **43**(2): 527-558.

Crivello, G. (2015). "'There's no future here': The time and place of children's migration aspirations in Peru." Geoforum **62**: 38-46.

Cross, C. (2001). "Why does South Africa need a spatial policy? Population migration, infrastructure and development." Journal of Contemporary African Studies **19**(1): 111-127.

Crowder, K. and S. South (2005). "Race, class, and changing patterns of migration between poor and nonpoor neighborhoods (vol 110, pg 1715, 2005)." American Journal of Sociology **111**(1): 350-355.

Crowder, K. and S. J. South (2008). "Spatial dynamics of white flight: The effects of local and extralocal racial conditions on neighborhood out-migration." American Sociological Review **73**(5): 792-812.

Crowder, K., S. J. South, et al. (2006). "Wealth, race, and inter-neighborhood migration." American Sociological Review **71**(1): 72-94.

Czaika, M. (2012). "Internal Versus International Migration and the Role of Multiple Deprivation: Some Evidence from India." Asian Population Studies **8**(2): 125-149.

Dabbaghian, V., P. Jackson, et al. (2010). "A cellular automata model on residential migration in response to neighborhood social dynamics." Mathematical and Computer Modelling **52**(9-10): 1752-1762.

Dai, H. J. (2011). "Surviving in "Localistic Communitas": Endogenous Multicultural Community Organizing among Migrant Workers in Post-Socialist China." Journal of Social Service Research **37**(2): 165-179.

Davin, D. (1996). "Migration and urbanisation in China - Day,LH, Xia,M." Pacific Review **9**(1): 131-132.

Davis, K. F., P. D'Odorico, et al. (2013). "Global spatio-temporal patterns in human migration: a complex network perspective." PLoS One **8**(1): e53723.

De Haan, A. (1999). "Livelihoods and poverty: The role of migration ‐ a critical review of the migration literature." The journal of development studies **36**(2): 1-47.

De Haas, H. (2007a). "Morocco's Migration Experience: A Transitional Perspective." International Migration **45**(4): 39-70.

De Haas, H. (2007b). "Turning the tide? Why development will not stop migration." Development and Change **38**(5): 819-841.

De Haas, H. (2009). "International Migration and Regional Development in Morocco: A Review." Journal of Ethnic and Migration Studies **35**(10): 1571-1593.

De Haas, H. (2010a). "The internal dynamics of migration processes: a theoretical inquiry." Journal of ethnic and migration studies **36**(10): 1587-1617.

De Haas, H. (2010b). "Migration and Development: A Theoretical Perspective." International Migration Review **44**(1): 227-264.

De Haas, H. (2011). "The determinants of international migration."

De Jong, G. F. and J. T. Fawcett (1981). "Motivations for migration: an assessment and a value-expectancy research model."

De Jong, G. F. and R. W. Gardner (2013). Migration decision making: multidisciplinary approaches to microlevel studies in developed and developing countries, Elsevier.

Démurger, S., J. D. Sachs, et al. (2002). "Geography, economic policy, and regional development in China." Asian Economic Papers **1**(1): 146-197.

Dennett, A. and A. Wilson (2013). "A multilevel spatial interaction modelling framework for estimating interregional migration in Europe." Environment and Planning A **45**(6): 1491-1507.

Deshingkar, P. (2006). "Internal migration, poverty and development in Asia: Including the excluded." Ids Bulletin-Institute of Development Studies **37**(3): 88-+.
Dorigo, G. and W. Tobler (1983). "Push-Pull Migration Laws." Annals of the Association of American Geographers **73**(1): 1-17.
Draper, N. and H. Smith (1966). "Applied regression analysis, New York, 1966." Google Scholar.
Du, J. G. and F. X. Cheng (2008). "A Study on Evolution of Rural-urban Income Inequality in China: 1993-2005." Recent Advance in Statistics Application and Related Areas, Pts 1 and 2: 1649-1654.
Ebenstein, A. and Y. H. Zhao (2015). "Tracking rural-to-urban migration in China: Lessons from the 2005 inter-census population survey." Population Studies-a Journal of Demography **69**(3): 337-353.
Eccles, J. S. (1987). "Gender roles and women's achievement-related decisions." Psychology of women Quarterly **11**(2): 135-172.
Eklund, L. (2000). "Gender Roles and Female Labour Migration—A Qualitative Field Study of Female Migrant Workers in Beijing." PROP Report.
Eldridge, J. D. and J. P. Jones III (1991). "Warped space: A geography of distance decay." The Professional Geographer **43**(4): 500-511.
Eng, I. (1997). "The rise of manufacturing towns: externally driven industrialization and urban development in the Pearl River Delta of China." International Journal of Urban and Regional Research **21**(4): 554-568.
Engen, S. (2007). "Stochastic growth and extinction in a spatial geometric Brownian population model with migration and correlated noise." Mathematical Biosciences **209**(1): 240-255.
Enyedi, G. (1992). "Urbanisation in East Central Europe: social processes and societal responses in the State Socialist Systems." Urban Studies **29**(6): 869-880.
Erdal, M. B. and C. Oeppen (2018). "Forced to leave? The discursive and analytical significance of describing migration as forced and voluntary." Journal of ethnic and migration studies **44**(6): 981-998.
Erikson, E. H. (1994). Identity: Youth and crisis, WW Norton & Company.
Evans, A. W. (1971). "The calibration of trip distribution models with exponential or similar cost functions." Transportation Research **5**: 15-38.
Evans, A. W. (1990). "The Assumption of Equilibrium in the Analysis of Migration and Interregional Differences - a Review of Some Recent Research." Journal of Regional Science **30**(4): 515-531.
Ezra, M. and G. E. Kiros (2001). "Rural out-migration in the drought prone areas of Ethiopia: A multilevel analysis." International Migration Review **35**(3): 749-771.
Fan, C. C. (1995). "Of Belts and Ladders - State Policy and Uneven Regional-Development in Post-Mao China." Annals of the Association of American Geographers **85**(3): 421-449.
Fan, C. C. (1996). "Economic opportunities and internal migration: A case study of Guangdong Province, China." Professional Geographer **48**(1): 28-45.
Fan, C. C. (1997). "Uneven development and beyond: Regional development theory in post-Mao China." International Journal of Urban and Regional Research **21**(4): 620-+.
Fan, C. C. (1999). "Migration in a socialist transitional economy: Heterogeneity, socioeconomic and spatial characteristics of migrants in China and Guangdong province." International Migration Review **33**(4): 954-987.
Fan, C. C. (2001). "Migration and labor-market returns in urban China: Results from a recent survey in Guangzhou." Environment and Planning A **33**(3): 479-508.
Fan, C. C. (2002). "The elite, the natives, and the outsiders: Migration and labor market segmentation in urban China." Annals of the Association of American Geographers **92**(1): 103-124.
Fan, C. C. (2003). "Rural ‐ urban migration and gender division of labor in transitional China." International Journal of Urban and Regional Research **27**(1): 24-47.
Fan, C. C. (2005a). "Interprovincial migration, population redistribution, and regional development in China: 1990 and 2000 census comparisons." The Professional Geographer **57**(2): 295-311.
Fan, C. C. (2005b). "Modeling interprovincial migration in China, 1985-2000." Eurasian Geography and Economics **46**(3): 165-184.
Fan, C. C. (2007). China on the Move: Migration, the State, and the Household, Routledge.

Fan, C. C. and Y. Huang (1998). "Waves of rural brides: Female marriage migration in China." Annals of the Association of American Geographers **88**(2): 227-251.
Fan, C. C. and L. Li (2002). "Marriage and migration in transitional China: a field study of Gaozhou, western Guangdong." Environment and Planning A **34**(4): 619-638.
Fan, C. C. and M. Sun (2008). "Regional inequality in China, 1978-2006." Eurasian geography and Economics **49**(1): 1-18.
Fan, C. C., M. Sun, et al. (2011). "Migration and split households: a comparison of sole, couple, and family migrants in Beijing, China." Environment and Planning-Part A **43**(9): 2164.
Fan, C. C. and W. W. Wang (2008). "The household as security: Strategies of rural–urban migrants in China." Migration and social protection in China **14**: 205.
Fan, S., R. Kanbur, et al. (2009). Regional inequality in China: trends, explanations and policy responses, Routledge.
Fan, S., R. Kanbur, et al. (2011). "China's regional disparities: Experience and policy." Review of Development Finance **1**(1): 47-56.
Featherstone, M. (1988). "In pursuit of the postmodern: an introduction." Theory, culture & society **5**(2-3): 195-215.
Feng, Y. (2002). "Population Growth, Urbanisation and the Role of Government in China: A Political Economic Model of Demographic Change." Urban studies **39**(12): 2329-2343.
Ferber, M. A. and B. G. Birnbaum (1977). "The "new home economics:" Retrospects and prospects." Journal of Consumer Research **4**(1): 19-28.
Field, A. (2013). Discovering statistics using IBM SPSS statistics, sage.
Fielding, A. J. (1982). "Counterurbanisation in western Europe." Progress in planning **17**: 1-52.
Fielding, A. J. (1992). "Migration and social mobility: South East England as an escalator region." Regional studies **26**(1): 1-15.
Fielding, T. (2010). Inter-regional migration in a transition economy: The case of China, Working Paper.
Fields, G. S. (1994). "The migration transition in Asia." Asian and Pacific Migration Journal **3**(1): 7-30.
Findley, S. E. (1977). Planning for internal migration: a review of issues and policies in developing countries, US Dept. of Commerce, Bureau of the Census, International Statistical Programs Center: for sale by the Supt. of Docs., US Govt. Print. Off.
Fingleton, B. (1986). "Analyzing Cross-Classified Data with Inherent Spatial Dependence." Geographical Analysis **18**(1): 48-61.
Finney, N. and L. Simpson (2009). 'Sleepwalking to segregation'?: challenging myths about race and migration, Policy Press.
Fitzgerald, J., D. Leblang, et al. (2014). "Defying the Law of Gravity: The Political Economy of International Migration." World Politics **66**(3): 1-49.
Fitzgerald, S., X. Chen, et al. (2013). "Occupational injury among migrant workers in China: a systematic review." Injury Prevention **19**(5): 348-354.
Fleisher, B., H. Li, et al. (2010). "Human capital, economic growth, and regional inequality in China." Journal of development economics **92**(2): 215-231.
Flores, M., M. Zey, et al. (2013). "Economic Liberalization and Contemporary Determinants of Mexico's Internal Migration: An Application of Spatial Gravity Models." Spatial Economic Analysis **8**(2): 195-214.
Flowerdew, R. (1982). "Fitting the lognormal gravity model to heteroscedastic data." Geographical Analysis **14**(3): 263-267.
Flowerdew, R. (2010). Modelling migration with Poisson regression. Technologies for migration and commuting analysis: Spatial interaction data applications, IGI Global**:** 261-279.
Flowerdew, R. and M. Aitkin (1982). "A method of fitting the gravity model based on the Poisson distribution." Journal of regional science **22**(2): 191-202.
Flowerdew, R. and J. Salt (1979). "Migration between labour market areas in Great Britain, 1970–1971." Regional Studies **13**(2): 211-231.
Fox, J. and S. Weisberg (2011). "Multivariate linear models in R." An R Companion to Applied Regression. Los Angeles: Thousand Oaks.
Francois, J. F. (1994). "Global Production and Trade - Factor Migration and Commercial-Policy with International Scale Economies." International Economic Review **35**(3): 565-581.

Froebel, F. (1980). The current development of the world economy, United Nations University Tokyo.

Froebel, F., J. Heinrichs, et al. (1980). The New Industrial Division of Labour: Structural Unemployment in Industrialized Countries and Industrialization in Developing Countries, Cambridge: Cambridge University Press.

Fu, Y. M. and S. A. Gabriel (2012). "Labor migration, human capital agglomeration and regional development in China." Regional Science and Urban Economics **42**(3): 473-484.

Fujita, M. and D. Hu (2001). "Regional disparity in China 1985–1994: The effects of globalization and economic liberalization." The Annals of Regional Science **35**(1): 3-37.

Fukuyama, F. (1995). "Social capital and the global economy." Foreign Aff. **74**: 89.

Gamlen, A. (2010). "The New Migration and Development Optimism: A Review of the 2009 Human Development Report." Global Governance **16**(3): 415-422.

Gao, S. (2012). "Commodification of place, consumption of identity: The sociolinguistic construction of a 'global village' in rural China." Journal of Sociolinguistics **16**(3): 336-357.

Gardner, R. W. (1981). "Macrolevel influences on the migration decision process."

Gedik, A. (2005). "Mobility transition (test of Zelinsky's theory) and economic and demographic factors: japan and turkey, 1955-2000."

Getis, A. (1990). "Screening for Spatial Dependence in Regression-Analysis." Papers of the Regional Science Association **69**: 69-81.

Ginsberg, R. B. (1972). "Incorporating Causal Structure and Exogenous Information with Probabilistic Models - with Special Reference to Choice, Gravity, Migration, and Markov Chains." Journal of Mathematical Sociology **2**(1): 83-103.

Giralt, R. M. (2015). "International Migration, Development and Human Wellbeing." Bulletin of Latin American Research **34**(3): 424-U178.

Gleditsch, K. S. and M. D. Ward (2001). "Measuring space: A minimum-distance database and applications to international studies." Journal of Peace Research **38**(6): 739-758.

Gleditsch, K. S. and M. D. Ward (2006). "Diffusion and the international context of democratization." International organization **60**(4): 911-933.

Gobermeyers, P. (1978). "Migration Analysis - Role of Geographic Scale." Annals of Regional Science **12**(3): 52-61.

Goldstein, A. and S. Goldstein (1987). "Migration in China - Methodological and Policy Challenges." Social Science History **11**(1): 85-104.

Goldstein, H. (2011). Multilevel statistical models, John Wiley & Sons.

Gong, X. and T. Huybers (2015). "Chinese students and higher education destinations: Findings from a choice experiment." Australian Journal of Education **59**(2): 196-218.

Gonzalez, R. and D. Griffin (1997). "On the statistics of interdependence: Treating dyadic data with respect." Handbook of personal relationships **2**: 271-301.

Gonzalez, R. and D. Griffin (2012). "Dyadic data analysis." APA handbook of research methods in psychology **3**: 439-450.

Goodman, J. L. (1981). "Information uncertainty and the microeconomic model of migration decision making."

Gordon, I. (1995). "Migration in a Segmented Labour Market." Transactions of the Institute of British Geographers **20**(2): 139-155.

Gray, C. L. and V. Mueller (2012). "Natural disasters and population mobility in Bangladesh." Proceedings of the National Academy of Sciences **109**(16): 6000-6005.

Greenhalgh, S. (2008). Just one child: Science and policy in Deng's China, Univ of California Press.

Greenwood, M. J. (1985). "Human migration: Theory, models, and empirical studies." Journal of regional Science **25**(4): 521-544.

Greenwood, M. J. (1993). "Migration - a Review." Regional Studies **27**(4): 295-296.

Greenwood, M. J. (2015). "International Handbook on Migration and Economic Development." Journal of Regional Science **55**(5): 880-883.

Greenwood, M. J. and G. L. Hunt (1989). "Jobs versus amenities in the analysis of metropolitan migration." Journal of Urban Economics **25**(1): 1-16.

Greenwood, M. J. J. (1985). "Human migration: theory, models, and empirical studies." Journal of regional science **25**(4): 521-544.

Gries, T., M. Kraft, et al. (2015). "Explaining inter-provincial migration in China." Papers in Regional Science: n/a-n/a.

Griffiths, M., A. Rogers, et al. (2013). "Migration, time and temporalities: review and prospect." COMPAS research resources paper, Oxford: Centre on Migration, Policy and Society, available at: www. compas. ox. ac. uk/publications/research-resources.

Grinblatt, M. and M. Keloharju (2001). "How distance, language, and culture influence stockholdings and trades." The Journal of Finance **56**(3): 1053-1073.

Groenewold, N., A. P. Chen, et al. (2010). "Inter-regional Spillovers of Policy Shocks in China." Regional Studies **44**(1): 87-101.

Gu, K. and G. Wall (2007). "Rapid urbanization in a transitional economy in China: The case of Hainan Island." Singapore Journal of Tropical Geography **28**(2): 158-170.

Gui, Y. X., J. W. Berry, et al. (2012). "Migrant worker acculturation in China." International Journal of Intercultural Relations **36**(4): 598-610.

Guo, F. and R. Iredale (2004). "The Impact of Hukou Status on Migrants' Employment: Findings from the 1997 Beijing Migrant Census 1." International Migration Review **38**(2): 709-731.

Guy, C. M. (1987). "Recent advances in spatial interaction modelling: an application to the forecasting of shopping travel." Environment and Planning A **19**(2): 173-186.

Hahn, H. P. (2010). "Urban Life-Worlds in Motion: In Africa and Beyond." Africa Spectrum **45**(3): 115-129.

Han, J., Y. Hayashi, et al. (2009). "Application of an integrated system dynamics and cellular automata model for urban growth assessment: A case study of Shanghai, China." Landscape and Urban Planning **91**(3): 133-141.

Han, S. S. and C. W. Pannell (1999). "The geography of privatization in China, 1978–1996." Economic Geography **75**(3): 272-296.

Han, S. S. and Z. X. Yan (1999). "China's coastal cities: Development, planning and challenges." Habitat International **23**(2): 217-229.

Han, X. R. and Y. F. Chen (2016). "Food consumption of outgoing rural migrant workers in urban area of China A QUAIDS approach." China Agricultural Economic Review **8**(2): 230-249.

Harris, J. R. and M. P. Todaro (1970). "Migration, unemployment and development: a two-sector analysis." The American economic review **60**(1): 126-142.

Harris, N. (1990). "Urbanisation, economic development and policy in developing countries." Habitat International **14**(4): 3-42.

Harris, R. (2017). Fitting a multilevel index of segregation in R. using the MLID package.

He, J. S. and J. Pooler (2002). "The regional concentration of China's interprovincial migration flows, 1982-90." Population and Environment **24**(2): 149-182.

He, S., M. M. Bayrak, et al. (2017). "A comparative analysis of multi-scalar regional inequality in China." Geoforum **78**(Supplement C): 1-11.

Head, K. and J. Ries (1996). "Inter-city competition for foreign investment: static and dynamic effects of China's incentive areas." Journal of Urban Economics **40**(1): 38-60.

Healy, K. (2005). "Book review: An R and S-plus companion to applied regression." Sociological Methods & Research **34**(1): 137-140.

Hendrischke, H. J. and C. Feng (1999). The political economy of China's provinces: comparative and competitive advantage, Psychology press.

Herrerias, M. J. (2012). "Weighted convergence and regional growth in China: an alternative approach (1952-2008)." Annals of Regional Science **49**(3): 685-718.

Holm, S. (1979). "A simple sequentially rejective multiple test procedure." Scandinavian journal of statistics: 65-70.

Howell, A. (2011). "Labor Market Segmentation in Urumqi, Xinjiang: Exposing Labor Market Segments and Testing the Relationship between Migration and Segmentation." Growth and Change **42**(2): 200-226.

Howell, A. and C. C. Fan (2011). "Migration and inequality in xinjiang: a survey of Han and Uyghur migrants in Urumqi." Eurasian Geography and Economics **52**(1): 119-139.

Huang, J. P., C. Li, et al. (2015). "Least-squares reverse time migration with static plane-wave encoding." Chinese Journal of Geophysics-Chinese Edition **58**(6): 2046-2056.

Huang, P. and S. Zhan (2005). Internal migration in China: Linking it to development, International Organization for Migration.

Huang, S. L., X. H. Lu, et al. (2011). "Spatial Inequality of Rural to Urban Migration in China: Based on 2005 National 1% Population Sample Survey." Proceedings of 2011 International Conference on Construction and Real Estate Management, Vols 1 and 2: 817-823.

Huitema, B. E. and J. W. McKean (1991). "Autocorrelation estimation and inference with small samples." Psychological Bulletin **110**(2): 291.

Hunter, A. (2011). "Theory and Practice of Return Migration at Retirement: the Case of Migrant Worker Hostel Residents in France." Population Space and Place **17**(2): 179-192.

Huo, J., X. M. Wang, et al. (2016). "Statistical characteristics of dynamics for population migration driven by the economic interests." Physica a-Statistical Mechanics and Its Applications **451**: 123-134.

Hyman, G. M. (1969). "The calibration of trip distribution models." Environment and Planning A **1**(1): 105-112.

Ikerd, J. (2005). "Sustainable capitalism." A Matter of Common Sense.

Inoue, Y., A. Stickley, et al. (2016). "The Association Between Economic Development, Lifestyle Differentiation, and C-Reactive Protein Concentration Within Rural Communities in Hainan Island, China." American Journal of Human Biology **28**(2): 186-196.

Jargowsky, P. A. (2009). "Immigrants and neighbourhoods of concentrated poverty: assimilation or stagnation?" Journal of Ethnic and Migration Studies **35**(7): 1129-1151.

Jaynes, E. T. (1957). "Information theory and statistical mechanics." Physical review **106**(4): 620.

Jenicek, V. (2010). "World population - development, transition." Agricultural Economics-Zemedelska Ekonomika **56**(1): 1-15.

Jenks, G. F. and F. C. Caspall (1971). "Error on choroplethic maps: definition, measurement, reduction." Annals of the Association of American Geographers **61**(2): 217-244.

Jenks, M. and N. Dempsey (2007). "Defining the neighbourhood: challenges for empirical research." Town Planning Review **78**(2): 153-177.

Jia, W. and X. Y. Liu (2014). "How much did the return of rural migrant labor affect China's national economy?" China Agricultural Economic Review **6**(1): 38-54.

Jian, T., J. D. Sachs, et al. (1996). "Trends in regional inequality in China." China economic review **7**(1): 1-21.

Jiang, C., J. Wang, et al. (2013). "Population Spatial Migration Tendency Forecasting with Grey Model and Fourier Series." Sustainable Development and Environment Ii, Pts 1 and 2 **409-410**: 69-74.

Jiang, P. (2004). "The relationship between ownership structure and firm performance: An empirical analysis over heilongjiang listed companies." Nature and Science **2**(4): 87-90.

Jivraj, S. (2012). "Modelling Socioeconomic Neighbourhood Change due to Internal Migration in England." Urban Studies **49**(16): 3565-3578.

Jivraj, S., M. Brown, et al. (2013). "Modelling Spatial Variation in the Determinants of Neighbourhood Family Migration in England with Geographically Weighted Regression." Applied Spatial Analysis and Policy **6**(4): 285-304.

Johnson, E. A. J. (1970). "The organization of space in developing countries." The organization of space in developing countries.

Johnston, R., D. Owen, et al. (2016). "House price increases and higher density housing occupation: the response of non-white households in London, 2001–2011." International Journal of Housing Policy **16**(3): 357-375.

Johnston, R. J. (1970). "Latent Migration Potential and Gravity Model - New-Zealand Study." Geographical Analysis **2**(4): 387-397.

Kaida, N. and T. M. Miah (2015). "Rural-urban perspectives on impoverishment risks in development-induced involuntary resettlement in Bangladesh." Habitat International **50**: 73-79.

Kallan, J. E. (1993). "A Multilevel Analysis of Elderly Migration." Social Science Quarterly **74**(2): 403-416.

Kanbur, R. and X. Zhang (1999). "Which regional inequality? The evolution of rural–urban and inland–coastal inequality in China from 1983 to 1995." Journal of comparative economics **27**(4): 686-701.

Kang, G. Y., J. M. Xiao, et al. (2016). "Congenital Heart Disease in Local and Migrant Elementary Schoolchildren in Dongguan, China." American Journal of Cardiology **117**(3): 461-464.

Keidel, A. (2009). "Chinese Regional Inequalities in Income and Well-Being." Review of Income and Wealth **55**: 538-561.

Keller, J. M., M. R. Gray, et al. (1985). "A fuzzy k-nearest neighbor algorithm." IEEE transactions on systems, man, and cybernetics(4): 580-585.

Kenny, D. A. and D. A. Kashy (2011). "Dyadic data analysis using multilevel modeling." Handbook of advanced multilevel analysis: 335-370.

Keung Wong, D. F., C. Y. Li, et al. (2007). "Rural migrant workers in urban China: living a marginalised life." International Journal of Social Welfare **16**(1): 32-40.

Khoo, S. E., G. Hugo, et al. (2008). "Which Skilled Temporary Migrants Become Permanent Residents and Why? 1." International Migration Review **42**(1): 193-226.

King, R. (2011). "Geography and migration studies: retrospect and prospect." Population space and place **18**(2): 134-153.

King, R. (2012). "Theories and typologies of migration: an overview and a primer."

King, R. and R. Skeldon (2010). "'Mind the gap!'Integrating approaches to internal and international migration." Journal of Ethnic and Migration Studies **36**(10): 1619-1646.

Kirk, D. (1996). "Demographic transition theory." Population studies **50**(3): 361-387.

Kissinger, H. and N. Hormann (2011). On China, Penguin Press New York.

Kloos, H., R. Correa-Oliveira, et al. (2010). "The role of population movement in the epidemiology and control of schistosomiasis in Brazil: a preliminary typology of population movement." Memorias Do Instituto Oswaldo Cruz **105**(4): 578-586.

Knight, A. P. and S. A. Humphrey "Dyadic Data Analysis."

Knight, J. and L. Song (1999). "The rural-urban divide: Economic disparities and interactions in China." OUP Catalogue.

Kofman, E. (2004). "Family-related migration: A critial review of European studies." Journal of Ethnic and Migration Studies **30**(2): 243-262.

Kontuly, T. and H. S. Geyer (2003). "Introduction to special issue: testing the differential urbanisation model in developed and less developed countries." Tijdschrift voor economische en sociale geografie **94**(1): 3-10.

Koster, J. M. and G. Leckie (2014). "Food sharing networks in lowland Nicaragua: an application of the social relations model to count data." Social Networks **38**: 100-110.

Kostov, P. (2010). "Model boosting for spatial weighting matrix selection in spatial lag models." Environment and Planning B: Planning and Design **37**(3): 533-549.

Kuang, L. and L. Liu (2012). "Discrimination against Rural-to-Urban Migrants: The Role of the Hukou System in China." Plos One **7**(11).

Kulakov, V., T. Sokur, et al. (1993). "Female reproductive function in areas affected by radiation after the Chernobyl power station accident." Environmental health perspectives **101**(Suppl 2): 117.

Kundu, A. (2003). "Urbanisation and urban governance: search for a perspective beyond neo-liberalism." Economic and political Weekly: 3079-3087.

Kundu, A. and S. Gupta (1996). "Migration, urbanisation and regional inequality." Economic and Political Weekly **31**(52): 3391-3398.

Laczko, F., J. Tjaden, et al. (2017). "Measuring Global Migration Potential, 2010–2015." Global Migration Data Briefing Series.

Lai, L. L. (2014). "Guest Editor's Introduction: The Local Intimacies of China's Rural-Urban Divide." Positions-Asia Critique **22**(3): 543-550.

Leamer, E. E. (2012). The craft of economics: lessons from the Heckscher-Ohlin framework, MIT Press.

Leckie, G. (2013). Cross-Classified Multilevel Models-Concepts. LEMMA VLE Module 12, 1-60.

Leckie, G. and C. Charlton (2013). "Runmlwin-a program to Run the MLwiN multilevel modelling software from within stata." Journal of Statistical Software **52**(11): 1-40.

Lee, C. K. (2003). Pathways of labour insurgency. Chinese Society, Routledge**:** 87-109.

Lee, C. K. (2007). Against the law: Labor protests in China's rustbelt and sunbelt, Univ of California Press.

Lee, E. S. (1966). "A theory of migration." Demography **3**(1): 47-57.

Lee, J., J. Carling, et al. (2014). "The International Migration Review at 50: Reflecting on Half a Century of International Migration Research and Looking Ahead." International Migration Review **48**: S3-S36.

Lee, T. (1968). "Urban Neighbourhood as a Socio-Spatial Schema." Human Relations **21**(3): 241-267.

Lemoine, F., S. Poncet, et al. (2015). "Spatial rebalancing and industrial convergence in China." China Economic Review **34**: 39-63.

Leorato, S. and M. Mezzetti (2016). "Spatial Panel Data Model with Error Dependence: A Bayesian Separable Covariance Approach." Bayesian Analysis **11**(4): 1035-1069.

LeSage, J. P. (2008). "An introduction to spatial econometrics." Revue d'économie industrielle(3): 19-44.

Levy, M. (2010). "Scale-free human migration and the geography of social networks." Physica a-Statistical Mechanics and Its Applications **389**(21): 4913-4917.

Levy, M. B. and W. J. Wadycki (1974). "What Is Opportunity Cost of Moving - Reconsideration of Effects of Distance on Migration." Economic Development and Cultural Change **22**(2): 198-214.

Lewis, W. A. (1954a). "Economic development with unlimited supplies of labour." The manchester school **22**(2): 139-191.

Lewis, W. A. (1954b). "Unlimited Supplies of Labour." Manchester school.

Li, J. J. and D. M. Sologon (2014). "A Continuous Labour Supply Model in Microsimulation: A Life-Cycle Modelling Approach with Heterogeneity and Uncertainty Extension." Plos One **9**(11).

Li, L., Y. L. Lei, et al. (2017). "Crowding-out effect of coal industry investment in coal mining area: taking Shanxi province in China as a case." Environmental Science and Pollution Research **24**(29): 23290-23298.

Li, P. (2009). "Zhongguo chengxiang renkou qianyi de shikong tezheng jiqi yingxiang yinsu." Economist(1): 50-57.

Li, Q., N. Stoeckl, et al. (2017). "Exploring the impacts of coal mining on host communities in Shanxi, China - using subjective data." Resources Policy **53**: 125-134.

Li, S. m. (2004). "Population migration and urbanization in China: A comparative analysis of the 1990 population census and the 1995 national one percent sample population survey." International Migration Review **38**(2): 655-685.

Li, Y., H. Liu, et al. (2014). "Analysis of Determinants on China's Interprovincial Migration during 1985-2005." Sustainable Development of Industry and Economy, Pts 1 and 2 **869-870**: 1096-1105.

Liang, Z., Y. P. Chen, et al. (2002). "Rural industrialisation and internal migration in China." Urban Studies **39**(12): 2175-2187.

Liang, Z., Z. Li, et al. (2014). "Changing Patterns of the Floating Population in China, 2000–2010." Population and Development Review **40**(4): 695-716.

Liang, Z. and M. J. White (1997). "Market transition, government policies, and interprovincial migration in China: 1983-1988." Economic Development and Cultural Change **45**(2): 321-339.

Lin, G. C. S. (1994). "Changing theoretical perspectives on urbanisation in Asian developing countries." Third World Planning Review **16**(1): 1.

Lin, G. C. S. (2001). "Evolving spatial form of urban-rural interaction in the Pearl River Delta, China." Professional Geographer **53**(1): 56-70.

Lin, W. C. (2016). "The Taiwan Independence Movement In and Out of Power." China Quarterly **227**: 842-843.

Ling, M. H. (2015). ""Bad Students Go to Vocational Schools!": Education, Social Reproduction and Migrant Youth in Urban China." China Journal **73**: 108-131.

Liu, G. G. (2018). Urbanization and Social Welfare in China, Routledge.

Liu, L. (2015). Chinese Student Migration and Selective Citizenship: Mobility, Community and Identity Between China and the United States, Routledge.

Liu, R. L., Q. X. Dong, et al. (2009). "Population Migration Model with System Dynamics Simulation." Proceedings of 2009 Conference on Systems Science, Management Science & System Dynamics, Vol 4: 161-164.

Liu, S. Q., F. T. Xie, et al. (2014). "Influences on rural migrant workers' selection of employment location in the mountainous and upland areas of Sichuan, China." Journal of Rural Studies **33**: 71-81.

Liu, T., Y. J. Qi, et al. (2015). "Spatial patterns, driving forces, and urbanization effects of China's internal migration: County-level analysis based on the 2000 and 2010 censuses." Journal of Geographical Sciences **25**(2): 236-256.

Liu, Y. (2012). "Does Internal Immigration Always Lead to Urban Unemployment in Emerging Economies?: A Structural Approach Based on Data from China." Hitotsubashi Journal of Economics **53**(1): 85-105.

Liu, Y. and J. Shen (2014a). "Jobs or amenities? Location choices of interprovincial skilled migrants in China, 2000–2005." Population, Space and Place **20**(7): 592-605.

Liu, Y. and J. Shen (2014b). "Spatial patterns and determinants of skilled internal migration in China, 2000–2005." Papers in Regional Science **93**(4): 749-771.

Liu, Y., J. Stillwell, et al. (2014). "Interprovincial Migration, Regional Development and State Policy in China, 1985–2010." Applied Spatial Analysis and Policy **7**(1): 47-70.

Liu, Y. Z. (2012). "Decisions to leave home: An examination of rural married women's labour migration in contemporary China." Womens Studies International Forum **35**(5): 305-313.

Liu, Z. (2005). "Institution and inequality: the hukou system in China." Journal of comparative economics **33**(1): 133-157.

Lomax, N., J. Stillwell, et al. (2014). "Internal Migration in the United Kingdom: Analysis of an Estimated Inter-District Time Series, 2001-2011." Applied Spatial Analysis and Policy **7**(1): 25-45.

Long, H. L. (2014). "Land consolidation: An indispensable way of spatial restructuring in rural China." Journal of Geographical Sciences **24**(2): 211-225.

Lozano-Ascencio, F., B. R. Roberts, et al. (1996). The interconnectedness of internal and international migration: the case of the United States and Mexico, Texas Population Research Center, The University of Texas at Austin.

Lu, M. and H. Gao (2011). "Labour market transition, income inequality and economic growth in China." International Labour Review **150**(1-2): 101-126.

Lu, M. and G. H. Wan (2014). "Urbanization and Urban Systems in the People's Republic of China: Research Findings and Policy Recommendations." Journal of Economic Surveys **28**(4): 671-685.

Lu, S., Y. T. Lin, et al. (2016). "Well-being of migrant and left-behind children in China: Education, health, parenting, and personal values." International Journal of Social Welfare **25**(1): 58-68.

Lu, Y. and F. Wang (2013). "From general discrimination to segmented inequality: Migration and inequality in urban China." Social science research **42**(6): 1443-1456.

Lu, Y. and H. Zhou (2013). "Academic Achievement and Loneliness of Migrant Children in China: School Segregation and Segmented Assimilation." Comparative Education Review **57**(1): 85-116.

Lucas, R. E. (1997). "Internal migration in developing countries." Handbook of population and family economics **1**: 721-798.

Lucas, R. E. B. (1978). "Internal Migration in Developing-Countries - Review of Theory, Evidence, Methodology and Research Priorities - Todaro,Mp." Journal of Development Economics **5**(1): 101-103.

Lucas, R. E. B. (2006). "Migration and economic development in Africa: A review of evidence." Journal of African Economies **15**: 337-395.

Luo, D. C., C. H. Gao, et al. (2011). "FDI, China's Non-agricultural Employment and Urban-Rural Income Inequality." Education and Management **210**: 247-253.

Luo, X., J. Shen, et al. (2014). "Urban and Regional Governance in China: Introduction." China Review **14**(1): 1-9.

Luo, X. B., N. Zhu, et al. (2014). "China's Lagging Region Development and Targeted Transportation Infrastructure Investments." Annals of Economics and Finance **15**(2): 365-409.

Lupton, R. (2003). "Neighbourhood effects: can we measure them and does it matter?".

Ma, L. J. (2002). "Urban transformation in China, 1949–2000: a review and research agenda." Environment and planning A **34**(9): 1545-1569.

Ma, L. J. and M. Fan (1994). "Urbanisation from below: The growth of towns in Jiangsu, China." Urban Studies **31**(10): 1625-1645.

Machado, C. C. and R. Hakkert (1988). "[An exploratory analysis of information on inter- and intra-municipal migration in the 1980 census]." Rev Bras Estud Popul **5**(2): 1-20.

Maoz, Z. (2004). "Pacifism and fightaholism in international politics: A structural history of national and dyadic conflict, 1816–1992." International Studies Review **6**(4): 107-133.

Massey, D. S., J. Arango, et al. (1993). "Theories of International Migration - a Review and Appraisal." Population and Development Review **19**(3): 431-466.

May, J. and N. Thrift (2003). Timespace: geographies of temporality, Routledge.

Maya-Jariego, I. and N. Armitage (2007). "Multiple senses of community in migration and commuting - The interplay between time, space and relations." International Sociology **22**(6): 743-766.

McKenzie, D. and D. Yang (2015). "Evidence on Policies to Increase the Development Impacts of International Migration." World Bank Research Observer **30**(2): 155-192.

Meentemeyer, V. (1989). "Geographical perspectives of space, time, and scale." Landscape ecology **3**(3-4): 163-173.

Meeus, B. (2012). "How to 'catch' floating populations? Research and the fixing of migration in space and time." Ethnic and Racial Studies **35**(10): 1775-1793.

Mendola, M. (2012). "Rural out-migration and economic development at origin: A review of the evidence." Journal of International Development **24**(1): 102-122.

Mi, H., W. Zhou, et al. (2009). "Renkou qianyi zhongli moxing de xiuzheng ji qi yingyong." Population Study **33**(4): 99-104.

Moeller, K. (1994). "Taiwan between Annexation and Independence." Aussen Politik **45**(2): 199-208.

Mohtadi, H. (1990). "Rural-Inequality and Rural-Push Versus Urban-Pull Migration - the Case of Iran, 1956-1976." World Development **18**(6): 837-844.

Molho, I. (2013). "Theories of Migration: A Review." Scottish Journal of Political Economy **60**(5): 526-556.

Møller, A. P. and T. A. Mousseau (2006). "Biological consequences of Chernobyl: 20 years on." Trends in ecology & evolution **21**(4): 200-207.

Moran, P. A. (1950). "Notes on continuous stochastic phenomena." Biometrika **37**(1/2): 17-23.

Moretti, E. (1999). "Social networks and migrations: Italy 1876-1913." International Migration Review: 640-657.

Mortimer, J. T. and M. J. Shanahan (2007). Handbook of the life course, Springer Science & Business Media.

Morton, N. (1977). "Isolation by distance in human populations." Annals of human genetics **40**(3): 361-365.

Myrdal, G. (1957). Rich lands and poor.

Nath, V. (1988). "Studies in Migration - Internal and International Migration in India - Rao,Msa." Economic and Political Weekly **23**(21): 1061-1066.

Nawrotzki, R. J. and M. Bakhtsiyarava (2017). "International climate migration: Evidence for the climate inhibitor mechanism and the agricultural pathway." Population, Space and Place **23**(4): e2033.

Naybor, D. (2018). The Impact of China's Changing Hukou on Women's Rural Land Rights. Environment, Politics, and Society, Emerald Publishing Limited**:** 119-132.

Nee, V., S. Opper, et al. (2007). "Developmental state and corporate governance in China." Management and Organization Review **3**(1): 19-53.

Newbold, K. B. (2005). "Spatial scale, return and onward migration, and the Long-Boertlein index of repeat migration." Papers in Regional Science **84**(2): 281-290.

Newman, M. E. and J. Park (2003). "Why social networks are different from other types of networks." Physical Review E **68**(3): 036122.

Ngok, K. (2012). "Serving Migrant Workers: A Challenging Public Service Issue in China." Australian Journal of Public Administration **71**(2): 178-190.

Nguyen, M. T. N. and C. Locke (2014). "Rural-urban migration in Vietnam and China: gendered householding, production of space and the state." Journal of Peasant Studies **41**(5): 855-876.

Nkrumah, K. (1967). "Neo-colonialism: The last stage of imperialism."

Oakley, A. (2016). Sex, gender and society, Routledge.

Olsson, G. (1965). "Distance and human interaction a migration study." Geografiska Annaler: Series B, Human Geography **47**(1): 3-43.

Otoiu, A. (2014). "Getting your Migration Analysis Together by Integrating Internal and International Migration." Procedia - Social and Behavioral Sciences **149**: 685-690.

Ozden, C., H. Rapoport, et al. (2015). "Migration and Development: Guest Editors' Preface." Scandinavian Journal of Economics **117**(2): 296-302.

Page, J. and S. Plaza (2006). "Migration remittances and development: A review of global evidence." Journal of African Economies **15**: 245-336.

Pannell, C. W. (2012). "China's New Leaders: Perspectives on Geographic Trends and Realities." Eurasian Geography and Economics **53**(6): 702-712.

Parida, J. K. and K. R. Raman (2018). "14 India: rising trends of international and internal migration." Handbook of Migration and Globalisation: 226.

Park, J. and K. Kim (2015). "Internal migration of the elderly in Korea: A multilevel logit analysis of their migration decision." Asian and Pacific Migration Journal **24**(2): 187-212.

Peeters, L. (2012). "Gravity and Spatial Structure: The Case of Interstate Migration in Mexico." Journal of Regional Science **52**(5): 819-856.

Petersen, W. (1958). "A general typology of migration." American Sociological Review **23**(3): 256-266.

Petras, E. M. (1981). "3: The Global Labor Market in the Modern World-Economy." International Migration Review **15**(1_suppl): 44-63.

Phelan, C., A. Nigam, et al. (2011). "Labor History Symposium: Rohini Hensman, Workers, Unions and Global Capitalism: Lessons from India." Labor History **52**(4): 535-562.

Phuong, N. H. and J. McPeak (2010). "Leaving or Staying: Inter-Provincial Migration in Vietnam." Asian and Pacific Migration Journal **19**(4): 473-500.

Pine, F. (2014). "Migration as Hope Space, Time, and Imagining the Future." Current Anthropology **55**: S95-S104.

Piotrowski, M. and Y. Y. Tong (2013). "Straddling Two Geographic Regions: the Impact of Place of Origin and Destination on Return Migration Intentions in China." Population Space and Place **19**(3): 329-349.

Plane, D. A. (1984). "A systemic demographic efficiency analysis of US interstate population exchange, 1935–1980." Economic Geography **60**(4): 294-312.

Plane, D. A. (1994). "The wax and wane of interstate migration patterns in the USA in the 1980s: a demographic effectiveness field perspective." Environment and Planning A **26**(10): 1545-1561.

Poncet, S. (2006). "Provincial migration dynamics in China: Borders, costs and economic motivations." Regional Science and Urban Economics **36**(3): 385-398.

Poon, J. P. H. and Q. Y. Shang (2012). "Regional Wage Inequality in China, 1996-2010." Eurasian Geography and Economics **53**(3): 338-355.

Poot, J. (1986). "A System Approach to Modelling the Inter‐Urban Exchange of Workers in New Zealand." Scottish Journal of Political Economy **33**(3): 249-274.

Poprawe, M. (2015). "On the relationship between corruption and migration: empirical evidence from a gravity model of migration." Public Choice **163**(3-4): 337-354.

Portes, A. (2007). "Migration, development, and segmented assimilation: A conceptual review of the evidence." Annals of the American Academy of Political and Social Science **610**: 73-97.

Posner, D. N. (2004). "The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi." American Political Science Review **98**(4): 529-545.

Potts, D. (2016). "Debates about African urbanisation, migration and economic growth: what can we learn from Zimbabwe and Zambia?" Geographical Journal **182**(3): 251-264.

Poveda, A. D. (2007). "Determinants and consequences of internal and international migration: The case of rural populations in the south of Veracruz, Mexico." Demographic Research **16**.

Prahalad, C. K. and K. Lieberthal (1998). "The end of corporate imperialism." Harvard Business Review **76**: 68-79.

Pumain, D. and L. Sanders (2013). "Theoretical Principles in Interurban Simulation Models: A Comparison." Environment and Planning A **45**(9): 2243-2260.

Qiu, P. Y., E. Caine, et al. (2011). "Depression and associated factors in internal migrant workers in China." Journal of Affective Disorders **134**(1-3): 198-207.

Qu, Z. F. and Z. Zhao (2014). "Evolution of the Chinese rural-urban migrant labor market from 2002 to 2007." China Agricultural Economic Review **6**(2): 316-334.

Rademacher-Schulz, C., B. Schraven, et al. (2014). "Time matters: shifting seasonal migration in Northern Ghana in response to rainfall variability and food insecurity." Climate and Development **6**(1): 46-52.

Radu, D. (2008). "Social interactions in economic models of migration: A review and appraisal." Journal of Ethnic and Migration Studies **34**(4): 531-548.

Raghuram, P. (2004). "The difference that skills make: gender, family migration strategies and regulated labour markets." Journal of ethnic and migration studies **30**(2): 303-321.

Rasbash, J., C. Charlton, et al. (2009). "MLwiN Version 2.1." Centre for multilevel modelling, University of Bristol.

Ravenstein, E. G. (1885). "The laws of migration." Journal of the statistical society of London **48**(2): 167-235.

Ravenstein, E. G. (1889). "The laws of migration." Journal of the royal statistical society **52**(2): 241-305.

Raymer, J., G. Abel, et al. (2007). "Combining census and registration data to estimate detailed elderly migration flows in England and Wales." Journal of the Royal Statistical Society: Series A (Statistics in Society) **170**(4): 891-908.

Raymer, J. and F. Wiilekens (2008). International migration in Europe: Data, models and estimates, John Wiley & Sons.

Reich, M., D. M. Gordon, et al. (1973). "A theory of labor market segmentation." The American Economic Review **63**(2): 359-365.

Relethford, J. H. (1986). "A Gravity Model of Human Migration." American Journal of Physical Anthropology **69**(2): 255-255.

Remoundou, K., M. Gkartzios, et al. (2016). "Conceptualizing Mobility in Times of Crisis: Towards Crisis-Led Counterurbanization?" Regional Studies **50**(10): 1663-1674.

Renaud, B. (1981). "National urbanization policy in developing countries."

Rezaei, S. (2015). "International Handbook on Migration and Economic Development." European Planning Studies **23**(11): 2327-2329.

Robertson, S. (2014). "Time and Temporary Migration: The Case of Temporary Graduate Workers and Working Holiday Makers in Australia." Journal of Ethnic and Migration Studies **40**(12): 1915-1933.

Rodgers, G. (1983). "The Impact of Rural-Urban Migration on Income-Distribution." Revue Canadienne D Etudes Du Developpement-Canadian Journal of Development Studies **4**(2): 275-296.

Rogers, A. and L. J. Castro (1981). "Model migration schedules."

Rogers, A., J. Raymer, et al. (2003). "Reconciling and translating migration data collected over time intervals of differing widths." Annals of Regional Science **37**(4): 581-601.

Rogers, A. and F. Willekens (1986). "Migration and settlement: A multiregional comparative study."

Rogerson, P. A., J. Raymer, et al. (1990). "Migration Analysis Using Data with Time Intervals of Differing Widths." Papers of the Regional Science Association **68**: 97-106.

Rossi, P. H. and A. B. Shlay (1982). "Residential mobility and public policy issues:“Why families move” revisited." Journal of Social Issues **38**(3): 21-34.

Rozelle, S., J. E. Taylor, et al. (1999). "Migration, remittances, and agricultural productivity in China." American Economic Review **89**(2): 287-291.

Rzhaniëtěsyna, L. S. (1983). Female labour under socialism, the socio-economic aspects, Progress Publishers (Moscow).

Saich, T. (2000). "Negotiating the state: The development of social organizations in China." The China Quarterly **161**: 124-141.

Schanbacher, G. L. (2007). Migration patterns : stories. Golden, Colo., Fulcrum Pub.

Schielzeth, H. (2010). "Simple means to improve the interpretability of regression coefficients." Methods in Ecology and Evolution **1**(2): 103-113.

Schiff, M. (2008). "On the underestimation of migration's income and poverty impact." Review of Economics of the Household **6**(3): 267-284.

Schiff, M. and C. Ozden (2007). International migration, economic development & policy, The World Bank.

Schiller, N. G. and A. Caglar (2009). "Towards a Comparative Theory of Locality in Migration Studies: Migrant Incorporation and City Scale." Journal of Ethnic and Migration Studies **35**(2): 177-202.

Schlichting, K., P. Tuckel, et al. (2015). "Great Migration of African Americans to Hartford, Connecticut, 1910-1930: A GIS Analysis at the Neighborhood and Street Level." Social Science History **39**(2): 287-310.

Schmeidl, S. (1997). "Exploring the causes of forced migration: A pooled time-series analysis, 1971-1990." Social Science Quarterly **78**(2): 284-308.

Schwind, P. J. (1973). "Comments on a Recent Review of Migration and Regional Development in United-States 1950-1960." Professional Geographer **25**(2): 174-175.

Serra, A. P. (2002). "Event study tests: a brief survey."

Seyfrit, C. L., T. Bjarnason, et al. (2010). "Migration Intentions of Rural Youth in Iceland: Can a Large-Scale Development Project Stem the Tide of Out-Migration?" Society & Natural Resources **23**(12): 1201-1215.

Shambaugh, D. L. (2013). China goes global: The partial power, Oxford University Press Oxford.

Shaw, R. P. (1975). "Migration theory and fact: a review and bibliography of current literature."

Shefer, D. and L. Steinvortz (1993). "Rural-to-urban and urban-to-urban migration patterns in Colombia." Habitat International **17**(1): 133-150.

Shen, J. (1999). "Modelling regional migration in China: estimation and decomposition." Environment and Planning A **31**(7): 1223-1238.

Shen, J. (2016a). "Error analysis of regional migration modeling." Annals of the American Association of Geographers: 1-15.

Shen, J. (2016b). "Modelling Interregional Migration in China in 2005–2010: the Roles of Regional Attributes and Spatial Interaction Effects in Modelling Error." Population, Space and Place.

Shen, J. and X. Yang (2014). "Analyzing Urban Competitiveness Changes in Major Chinese Cities 1995–2008." Applied Spatial Analysis and Policy **7**(4): 361-379.

Shen, J. F. (1996). "Internal migration and regional population dynamics in China." Progress in Planning **45**: 125-&.

Shen, J. F. (2012). "Changing Patterns and Determinants of Interprovincial Migration in China 1985-2000." Population Space and Place **18**(3): 384-402.

Shen, J. F. (2013). "Increasing internal migration in China from 1985 to 2005: Institutional versus economic drivers." Habitat International **39**: 1-7.

Shen, J. F. (2015). "Explaining Interregional Migration Changes in China, 1985-2000, Using a Decomposition Approach." Regional Studies **49**(7): 1176-1192.

Shen, J. F., Z. Q. Feng, et al. (2006). "Dual-track urbanization in a transitional economy: The case of Pearl River Delta in South China." Habitat International **30**(3): 690-705.

Sheng, Z. (2011). "Towards China's urban-rural integration: Issues and options." International Journal of China Studies **2**(2): 345-367.

Shenggen, F. and X. Zhang (2004). "Infrastructure and regional economic development in rural China." China economic review **15**(2): 203-214.

Shi, Y., X. Q. Zheng, et al. (2014). "Intra-urban Migration Patterns in the Perspective of Improved Radiation Model-a case study of Beijing." Sustainable Development of Industry and Economy, Pts 1 and 2 **869-870**: 110-116.

Shin, S. S., X. Wan, et al. (2013). "Perceived Discrimination and Smoking Among Rural-to-Urban Migrant Women in China." Journal of Immigrant and Minority Health **15**(1): 132-140.

Shishehgar, S., L. Gholizadeh, et al. (2015). "The impact of migration on the health status of Iranians: an integrative literature review." Bmc International Health and Human Rights **15**.

Sicular, T., Y. Ximing, et al. (2007). "The urban–rural income gap and inequality in China." Review of Income and Wealth **53**(1): 93-126.

Silvestre, J. and D. S. Reher (2014). "The Internal Migration of Immigrants: Differences between One-Time and Multiple Movers in Spain." Population Space and Place **20**(1): 50-65.

Simes, R. J. (1986). "An improved Bonferroni procedure for multiple tests of significance." Biometrika **73**(3): 751-754.
Sjaastad, L. A. (1962). "The costs and returns of human migration." Journal of political Economy **70**(5, Part 2): 80-93.
Sjoberg, O. (1999). "Shortage, priority and urban growth: Towards a theory of urbanisation under central planning." Urban Studies **36**(13): 2217-2236.
Skeldon, R. (1992). "On mobility and fertility transitions in East and Southeast Asia." Asian and Pacific Migration Journal **1**(2): 220-249.
Skeldon, R. (2006). "Interlinkages between internal and international migration and development in the Asian region." Population Space and Place **12**(1): 15-30.
Skeldon, R. (2014). Migration and development: A global perspective, Routledge.
Slater, P. B. (1985). "Point-to-Point Migration Functions and Gravity Model Renormalization - Approaches to Aggregation in Spatial Interaction Modeling." Environment and Planning A **17**(8): 1025-1044.
Smith-Greenaway, E. and K. J. A. Thomas (2014). "Exploring Child Mortality Risks Associated with Diverse Patterns of Maternal Migration in Haiti." Population Research and Policy Review **33**(6): 873-895.
Smith, P. W. F., J. Raymer, et al. (2010). "Combining available migration data in England to study economic activity flows over time." Journal of the Royal Statistical Society Series a-Statistics in Society **173**: 733-753.
Snyder, S. and W. S. Chern (2009). "The impact of remittance income on rural households in China." China Agricultural Economic Review **1**(1): 38-57.
Solis, P., T. W. Pullum, et al. (2007). "Homogamy by education and migration status in Monterrey, Mexico: changes and continuities over time." Population Research and Policy Review **26**(3): 279-298.
Solow, R. M. (1957). "Technical change and the aggregate production function." The review of Economics and Statistics **39**(3): 312-320.
Song, X. and Y. Wang (2005). "Zhongguo renkou kongjian fenbu bianhua de shoulianxing fenxi." Zhongguo renkou kexue(4): 54-58.
Stark, O. and O. Stark (1991). "The migration of labor."
Stark, O. and J. E. Taylor (1991). "Migration incentives, migration types: The role of relative deprivation." The economic journal **101**(408): 1163-1178.
StataCorp, L. (2007). "Stata data analysis and statistical Software." Special Edition Release **10**: 733.
Stewart, C. T. (1960). "Migration as a Function of Population and Distance." American Sociological Review **25**(3): 347-356.
Stouffer, S. A. (1940). "Intervening opportunities: a theory relating mobility and distance." American sociological review **5**(6): 845-867.
Streeck, W. (2014). "How will capitalism end?" New Left Review(87): 35-64.
Su, C. W., T. Y. Liu, et al. (2015). "Is urbanization narrowing the urban-rural income gap? A cross-regional study of China." Habitat International **48**: 79-86.
Sun, L. and T. Liu (2014). "Injured but not Entitled to Legal Insurance Compensation - Ornamental Institutions and Migrant Workers' Informal Channels in China." Social Policy & Administration **48**(7): 905-922.
Sun, S. P. and S. M. Manson (2012). "Intraurban Migration, Neighborhoods, and City Structure." Urban Geography **33**(7): 1008-1029.
Sun, X. X., C. G. Q. Chi, et al. (2013). "Developing Destination Loyalty: The Case of Hainan Island." Annals of Tourism Research **43**: 547-577.
Sun, Z. (2013). "Explaining Regional Disparities of China’s Economic Growth: Geography, Policy and Infrastructure." California: Universtiy of California, Berkeley.
Swain, A. (1996). "Environmental migration and conflict dynamics: focus on developing regions." Third World Quarterly **17**(5): 959-974.
Taaffe, E. J. (1996). Geography of transportation, Morton O'kelly.
Tao, L., E. C. M. Hui, et al. (2015). "Housing choices of migrant workers in China: Beyond the Hukou perspective." Habitat International **49**: 474-483.

Taylor, E. J. (1999). "The new economics of labour migration and the role of remittances in the migration process." International migration **37**(1): 63-88.

Taylor, J. E., J. Arango, et al. (1996). "International migration and community development." Population Index: 397-418.

Thacker, R. A. and G. R. Ferris (1991). "Understanding sexual harassment in the workplace: The influence of power and politics within the dyadic interaction of harasser and target." Human Resource Management Review **1**(1): 23-37.

Tharmaseelan, N., K. Inkson, et al. (2010). "Migration and career success: testing a time-sequenced model." Career Development International **15**(2-3): 218-238.

Thomas, M., J. Stillwell, et al. (2015). "Modelling multilevel variations in distance moved between origins and destinations in England and Wales." Environment and Planning A **47**(4): 996-1014.

Tian, Q. J. (2001). "China's new urban-rural divide and pitfalls for the Chinese economy." Revue Canadienne D Etudes Du Developpement-Canadian Journal of Development Studies **22**(1): 165-190.

Tobler, W. R. (1970). "A computer movie simulating urban growth in the Detroit region." Economic geography **46**(sup1): 234-240.

Todaro, M. (1980). Internal migration in developing countries: a survey. Population and economic change in developing countries, University of Chicago Press**:** 361-402.

Todaro Michael, P. (1976). "Internal migration in developing countries: A review of theory, evidence, methodology and research priorities." Geneva: International Labour Office.

Todaro, M. P. (1983). The struggle for economic development: readings in problems and policies, Pearson Education.

Todaro, M. P. and J. Stilkind (1981). "City bias and rural neglect: the dilemma of urban development."

Tsai, K. S. (2007). Capitalism without democracy: The private sector in contemporary China, Cornell University Press.

Van de Kaa, D. J. (1987). "Europe's second demographic transition." Population bulletin **42**(1): 1-59.

Van Hear, N., O. Bakewell, et al. (2018). "Push-pull plus: reconsidering the drivers of migration." Journal of ethnic and migration studies **44**(6): 927-944.

Vanderkamp, J. (1977). "Gravity Model and Migration Behavior - Economic Interpretation." Journal of Economic Studies **4**(2): 89-102.

VanWey, L. K. (2005). "Land ownership as a determinant of international and internal migration in Mexico and internal migration in Thailand." International Migration Review **39**(1): 141-172.

Vartiainen, P. (1989). "Counterurbanisation: a challenge for socio-theoretical geography." Journal of Rural Studies **5**(3): 217-225.

Vendryes, T. (2011). "Migration constraints and development: Hukou and capital accumulation in China." China Economic Review **22**(4): 669-692.

Vignoli, J. R. (2008). "Spatial distribution, internal migration and development in Latin America and the Caribbean." Cepal Review(96): 137-157.

Villarreal, A. and E. R. Hamilton (2012). "Rush to the border? Market liberalization and urban- and rural-origin internal migration in Mexico." Social Science Research **41**(5): 1275-1291.

Villaverde, J., A. Maza, et al. (2010). "Provincial Disparities in Post-reform China." China & World Economy **18**(2): 73-95.

Wahba, J. and Y. Zenou (2012). "Out of sight, out of mind: Migration, entrepreneurship and social capital." Regional Science and Urban Economics **42**(5): 890-903.

Wallerstein, I. (1974). "The rise and future demise of the world capitalist system: concepts for comparative analysis." Comparative studies in society and history **16**(4): 387-415.

Wan, G. H. (1995). "Peasant Flood in China - Internal Migration and Its Policy Determinants." Third World Quarterly **16**(2): 173-196.

Wang, C. (2001). "Social Identity of the New Generation of Rural Hobo and Merger of Urban and Rural [J]." Sociological Research **3**: 63-76.

Wang, C. B. (2013). "Place of desire: Skilled migration from mainland China to post-colonial Hong Kong." Asia Pacific Viewpoint **54**(3): 388-397.

Wang, F. and X. Zuo (1999). "Inside China's cities: Institutional barriers and opportunities for urban migrants." American Economic Review **89**(2): 276-280.

Wang, G. (1993). "Woguo shengji renkou qianyi yu juli guanxi zhi tantao." Population and Economy **2**: 3-8.
Wang, G. (2004). "The effects of regional income gap on migration: new empirical results from the 2000 census." China Economic Quarterly **3**(B10): 77-98.
Wang, G., X. Wei, et al. (2005). "Study on Inter-Provincial Migration's Influence on Regional Economy Development in China." Fudan Journal (Social Sciences) **3**: 148-161.
Wang, H. N., F. Guo, et al. (2015a). "Discrimination in Migrant Workers' Welfare Entitlements and Benefits in Urban Labour Market: Findings from a Four-City Study in China." Population Space and Place **21**(2): 124-139.
Wang, H. N., F. Guo, et al. (2015b). "A distributional analysis of wage discrimination against migrant workers in China's urban labour market." Urban Studies **52**(13): 2383-2403.
Wang, L. (2012). "Social exclusion and education inequality: towards an integrated analytical framework for the urban-rural divide in China." British Journal of Sociology of Education **33**(3): 409-430.
Wang, W., Q. X. Wu, et al. (2017). "Migrant elders in Hainan, China, report better health and lower depression than local-born elderly residents." Annals of Human Biology **44**(7): 622-627.
Wang, W. W. and C. C. Fan (2006). "Success or failure: selectivity and reasons of return migration in Sichuan and Anhui, China." Environment and Planning A **38**(5): 939-958.
Wang, W. W. and C. C. Fan (2012). "Migrant workers' integration in urban China: Experiences in employment, social adaptation, and self-identity." Eurasian Geography and Economics **53**(6): 731-749.
Wang, X. B. and J. Piesse (2010). "Inequality and the Urban-rural Divide in China: Effects of Regressive Taxation." China & World Economy **18**(6): 36-55.
Wang, X. B., J. Piesse, et al. (2013). "Mind the gaps: a political economy of the multiple dimensions of China's rural-urban divide." Asian-Pacific Economic Literature **27**(2): 52-67.
Wang, X. W., L. J. Zhang, et al. (2013). "Exact Solution of Population Redistributions in a Migration Model." International Journal of Modern Physics C **24**(10).
Wang, Y., Y. N. Chen, et al. (2013). "Evolvement Characteristics of Population and Economic Gravity Centers in Tarim River Basin, Uygur Autonomous Region of Xinjiang, China." Chinese Geographical Science **23**(6): 765-772.
Wang, Y. L., J. X. Hu, et al. (2016). "Health risk assessment of migrant workers' exposure to polychlorinated biphenyls in air and dust in an e-waste recycling area in China: Indication for a new wealth gap in environmental rights." Environment International **87**: 33-41.
Warner, M. (2011). "Hainan-state, society, and business in a Chinese province." Asia Pacific Business Review **17**(4): 548-549.
Wei, Y. H. D. (2007). "Regional development in China: Transitional institutions, embedded globalization, and hybrid economies." Eurasian Geography and Economics **48**(1): 16-36.
Wen, M. and D. H. Lin (2012). "Child Development in Rural China: Children Left Behind by Their Migrant Parents and Children of Nonmigrant Families." Child Development **83**(1): 120-136.
Widdis, R. W. (1988). "Scale and Context - Approaches to the Study of Canadian Migration Patterns in the 19th-Century." Social Science History **12**(3): 269-303.
Wilber, G. L. (1965). "A Bayesian Model for Migration Decisions in a Population." Population Index **31**(3): 257-257.
Willekens, F. (1983). "Log‐linear modelling of spatial interaction." Papers in Regional Science **52**(1): 187-205.
Willekens, F. (1994). "Monitoring international migration flows in Europe." European Journal of Population/Revue européenne de Démographie **10**(1): 1-42.
Willekens, F. (1999). "Modeling approaches to the indirect estimation of migration flows: From entropy to EM." Mathematical Population Studies **7**(3): 239-278.
Williamson, O. E. (1981). "The economics of organization: The transaction cost approach." American journal of sociology **87**(3): 548-577.
Wilson, A. G. (1969). "The use of entropy maximising models, in the theory of trip distribution, mode split and route split." Journal of Transport Economics and Policy: 108-126.
Wilson, A. G. (1971). "A family of spatial interaction models, and associated developments." Environment and Planning A **3**(1): 1-32.

Wilson, S. J. (2003). "A dynamic general equilibrium analysis of migration and capital formation: the case of Canada." Review of Economic Dynamics **6**(2): 455-481.
Wolpert, J. (1966). "Migration as an adjustment to environmental stress." Journal of Social Issues **22**(4): 92-102.
Wong, D. F. K., X. S. He, et al. (2008). "Mental health of migrant workers in China: prevalence and correlates." Social Psychiatry and Psychiatric Epidemiology **43**(6): 483-489.
Wood, C. H. (1982). "Equilibrium and historical-structural perspectives on migration." International Migration Review **16**(2): 298-319.
Wu, Q. and H. Xiao (2014). "Dynamic CGE Model and Simulation Analysis on the Impact of Citizenization of Rural Migrant Workers on the Labor and Capital Markets in China." Discrete Dynamics in Nature and Society.
Wu, Z. M. and S. J. Yao (2003). "Intermigration and intramigration in China: A theoretical and empirical analysis." China Economic Review **14**(4): 371-385.
Xing, L., S. G. Fan, et al. (2009). "Community poverty and inequality in western China: A tale of three villages in Guizhou Province." China Economic Review **20**(2): 338-349.
Xu, Q. W., X. P. Guan, et al. (2011). "Welfare program participation among rural-to-urban migrant workers in China." International Journal of Social Welfare **20**(1): 10-21.
Yan, S. (2007). "Inter-provincial migration and its determinants in the 1990's China." Chinese Journal of Population Science **1**: 71-77.
Yan, X. H., S. Bauer, et al. (2014). "Farm Size, Land Reallocation, and Labour Migration in Rural China." Population Space and Place **20**(4): 303-315.
Yang, C. F., H. R. Han, et al. (2014). "Spatial Distribution of Migration and Economic Development: A Case Study of Sichuan Province, China." Sustainability **6**(10): 6509-6528.
Yang, H., F. He, et al. (2015). "Health-Related Lifestyle Behaviors among Male and Female Rural-to-Urban Migrant Workers in Shanghai, China." Plos One **10**(2).
Yang, H., X. Li, et al. (2005). "Workplace and HIV-related sexual behaviours and perceptions among female migrant workers." AIDS care **17**(7): 819-833.
Yang, J. H. (2013). "Social Exclusion and Young Rural-Urban Migrants' Integration into a Host Society in China." Annals of the American Academy of Political and Social Science **648**(1): 52-69.
Yang, X. and F. Guo (1999). "Gender differences in determinants of temporary labor migration in China: A multilevel analysis." International Migration Review: 929-953.
Yang, X. S. (1994). "Urban Temporary out-Migration under Economic-Reforms - Who Moves and for What Reasons." Population Research and Policy Review **13**(1): 83-100.
Yang, X. S. (2000). "Determinants of migration intentions in Hubei province, China: individual versus family migration." Environment and Planning A **32**(5): 769-787.
Yang, X. S. and F. Guo (1999). "Gender differences in determinants of temporary labor migration in China: A multilevel analysis." International Migration Review **33**(4): 929-953.
Yapa, L. and J. Wolpert (1971). "Time Paths of Migration Flows - Belgium, 1954-62." Geographical Analysis **3**(2): 157-164.
Ye, J., C. Wang, et al. (2013). "Internal migration and left-behind populations in China." Journal of Peasant Studies **40**(6): 1119-1146.
Yeoh, B. S. and S. Huang (1998). "Negotiating public space: Strategies and styles of migrant female domestic workers in Singapore." Urban studies **35**(3): 583-602.
Yeoh, B. S., S. Huang, et al. (1999). "Migrant female domestic workers: debating the economic, social and political impacts in Singapore." International Migration Review: 114-136.
Yiu, D. W., C. Lau, et al. (2007). "International venturing by emerging economy firms: The effects of firm capabilities, home country networks, and corporate entrepreneurship." Journal of international business studies **38**(4): 519-540.
Yorimitsu, M. (1985). "A Review on the Determinants of Migration." Hitotsubashi Journal of Social Studies **17**(1): 17-27.
You, H. Y. and X. F. Yang (2017). "Urban expansion in 30 megacities of China: categorizing the driving force profiles to inform the urbanization policy." Land Use Policy **68**: 531-551.
Young, J. (2013). China's hukou system: Markets, migrants and institutional change, Springer.
Yu, L., W. Xu, et al. (2014). "Geographic structure of the migrant labour market in China: A case study of four city-regions in Fujian." Canadian Studies in Population **41**(3-4): 22-48.

Zelinsky, W. (1971). "The hypothesis of the mobility transition." Geographical review: 219-249.
Zelinsky, W. (1979). "The demographic transition: changing patterns of migration."
Zhan, S. H. (2011). "What Determines Migrant Workers' Life Chances in Contemporary China? Hukou, Social Exclusion, and the Market." Modern China **37**(3): 243-285.
Zhang, C. H., Q. J. Gao, et al. (2013). "The Impact of Rural-urban Migration on Gender Relations in Chinese Households." Asian Journal of Womens Studies **19**(1): 41-66.
Zhang, G. S., J. Lue, et al. (2004). "The generalized entropy index model and the decomposition of regional income inequality in rural China during transition." Proceedings of the 2004 International Conference on Management Science & Engineering, Vols 1 and 2: 1679-1683.
Zhang, K. H. and S. F. Song (2003). "Rural-urban migration and urbanization in China: Evidence from time-series and cross-section analyses." China Economic Review **14**(4): 386-400.
Zhang, L. and L. Tao (2012). "Barriers to the acquisition of urban hukou in Chinese cities." Environment and Planning A **44**(12): 2883-2900.
Zhang, M. Q., C. J. Zhu, et al. (2014). "The Institution of Hukou-based Social Exclusion: A Unique Institution Reshaping the Characteristics of Contemporary Urban China." International Journal of Urban and Regional Research **38**(4): 1437-1457.
Zhang, N. N. (2013a). "Rural women migrant returnees in contemporary China." Journal of Peasant Studies **40**(1): 171-188.
Zhang, Q. F. (2013b). "Gender Disparities in Self-employment in Urban China's Market Transition: Income Inequality, Occupational Segregation and Mobility Processes." China Quarterly(215): 744-763.
Zhang, X. and R. Kanbur (2005). "Spatial inequality in education and health care in China." China Economic Review **16**(2): 189-204.
Zhang, X. N. (2016). "Analysis of urban-rural divide and interprovincial migration in China from 2000 to 2010 with gravity models." Quadernidel Premio «Giorgio Rota» **4**: 69-82.
Zhang, X. N., W. W. Wang, et al. (2018). "Analysing inter-provincial urban migration flows in China: A new multilevel gravity model approach." Migration Studies.
Zhao, Y. (1999). "Labor migration and earnings differences: the case of rural China." Economic Development and Cultural Change **47**(4): 767-782.
Zhao, Y. (2003). "The role of migrant networks in labor migration: The case of China." Contemporary Economic Policy **21**(4): 500-511.
Zhong, B. L., T. B. Liu, et al. (2015). "Prevalence and correlates of major depressive disorder among rural-to-urban migrant workers in Shenzhen, China." Journal of Affective Disorders **183**: 1-9.
Zhou, J. and G. S. Zhang (2008). "Empirical Analysis of the Impact of FDI on China's Urban-Rural Income Inequality*-based on the Study of Inter-Provincial Panel Data." Proceedings of 2008 Conference on Regional Economy and Sustainable Development: 39-43.
Zhu, C.-g., R.-h. Ma, et al. (2005). "The Spatial Structure of the Floating Population in Chinese cities." Human Geography **17**(1): 65-68.
Zhu, C. Y., Q. S. Geng, et al. (2013). "Quality of life in China rural-to-urban female migrant factory workers: a before-and-after study." Health and Quality of Life Outcomes **11**.
Zhu, J. M. (1998). "Rural out-migration in China: A multilevel model." Migration, Urbanization, and Development : New Directions and Issues: 157-186.
Zhu, N. (2002). "The impacts of income gaps on migration decisions in China." China Economic Review **13**(2-3): 213-230.
Zhu, Y. (2018). "Advancing research on internal migration in Asia: the mobility transition hypothesis revisited." Asian Population Studies **14**(1): 1-4.
Zhu, Y., M. Bell, et al. (2013). "Rural-Urban Linkages and the Impact of Internal Migration in Asian Developing Countries an Introduction." Asian Population Studies **9**(2): 119-123.
Zhu, Y., L. Lin, et al. (2016). "Trends in internal migration and mobility: International experiences and their implications for China." Renkou Yanjiu (Population Research) **40**(5): 50-60.
Zhuang, M. (2009). "The Social Support Network for Rural Migrant Workers in Chengdu, China: Local Governance and Civil Society in the Fight Against Poverty and Exclusion." Ids Bulletin-Institute of Development Studies **40**(6): 41-49.
Zhukov, Y. M. and B. M. Stewart (2013). "Choosing Your Neighbors: Networks of Diffusion in International Relations." International Studies Quarterly **57**(2): 271-287.

Zipf, G. K. (1946). "The P₁ P₂/D Hypothesis: On the Intercity Movement of Persons." American Sociological Review **11**(6): 677-686.

Zuo, J. (2008). "Marital construction of family power among male-out-migrant couples in a Chinese village - A relation-oriented exchange model." Journal of Family Issues **29**(5): 663-691.

Appendix

Chapter 1 Introduction

The cross-classified migration model with the two additional correlations in Chapter 5 is the core model of this thesis, with more complicated models in Chapter 6 expanding to include a few more variables based on it. Therefore only the syntax for it is included below to demonstrate the model set-up procedure, whilst syntaxes of models in Chapter 6 share the same multilevel structure so will not be further discussed.

```
* Load the data

use "O:\thesis\stata for thesis\10 ml migtype-residual.dta", clear

* Set MLwiN software pathway

global MLwiN_path C:\Program Files (x86)\MLwiN v2.36(BETA)\i386\mlwin.exe

set matsize 10000

* Keep variables of interest

keep china oid o1-o31 did d1-d31 pid sid z1 z2 ///

  lnmuu lnoupop lndupop lnouincome lnduincome lndistance cons

* Sort the data acording to the hierarchy implied by the model

sort china pid sid

* Define vector b which indicates which elements of the China matrix are to be

* freely estimated (=1) and which elements are to be constrained to zero (=0)

matrix a = I(31)

matrix a = (a,a \ a,a)
```

```
matrix list a
matrix b = 1
forvalues r = 2/62 {
  matrix b = (b,a[`r',1..`r'])
}
matrix list b
* Define constraints for the China matrix
constraint drop _all
forvalues c = 1/30 {
  constraint define `c' [RP3]var(o`=`c'+1') = [RP3]var(o1)
}
forvalues c = 31/60 {
  constraint define `c' [RP3]var(d`=`c'-30 + 1') = [RP3]var(d1)
}
forvalues c = 61/90 {
  constraint define `c' [RP3]cov(o`=`c'-60 + 1'\d`=`c'-60 + 1') = [RP3]cov(o1\d1)
}
constraint define 91 [RP2]var(z1) = [RP2]var(z2)
* Fit the model
runmlwin lnmuu lnoupop lndupop lnouincome lnduincome lndistance cons, ///
  level3(china: o1-o31 d1-d31, elements(b) residuals(uv)) ///
```

```
  level2(pid: z1 z2, residuals(e)) ///

  level1(sid: cons) ///

  constraints(1/91) ///

  mlwinsettings(variables(124)) ///

  nopause

* Format the predicted random effects and residuals

format %9.2f uv* e*

* Compress and save the data

compress

save "temp.dta", replace
```

Appendix

Chapter 3 Data and methodology

Table 3.6 Full interprovincial migration flow matrix in 2000 (Unit: 1000 persons)

		Origin																													
		Beijing	Tianjin	Hebei	Shanxi	Inner Mongolia	Liaoning	Jilin	Heilongjiang	Shanghai	Jiangsu	Zhejiang	Anhui	Fujian	Jiangxi	Shandong	Henan	Hubei	Hunan	Guangdong	Guangxi	Hainan	Chongqing	Sichuan	Guizhou	Yunnan	Tibet	Shaanxi	Gansu	Qinghai	Ningxia	Xinjiang
Destination	Beijing		18	555	66	68	55	47	90	4	141	93	228	45	47	189	335	105	37	23	8	2	21	168	13	5	1	49	28	4	5	11
	Tianjin	4		203	11	21	15	19	52	1	20	20	59	11	4	122	74	25	5	3	2	0	5	33	3	1	0	10	8	1	1	3
	Hebei	29	32		37	62	37	33	112	1	25	40	50	12	8	66	105	55	12	6	5	1	7	106	13	8	0	44	12	3	2	4
	Shanxi	2	1	85		100	5	4	4	1	15	30	28	11	3	25	115	30	8	2	2	0	6	109	9	8	0	49	9	2	1	1
	Inner Mongolia	2	3	76	44		35	33	118	0	14	15	12	4	4	22	21	6	3	1	0	0	2	26	1	1	0	47	40	2	14	1
	Liaoning	3	2	32	5	97		174	320	1	34	31	56	15	7	92	59	25	6	5	1	0	5	57	2	1	0	7	3	1	1	1
	Jilin	1	0	11	1	22	41		91	1	14	11	14	4	2	46	11	15	3	1	0	0	1	14	0	0	0	1	1	0	0	0
	Heilongjiang	1	1	15	1	35	39	99		1	26	12	22	4	3	77	15	14	3	2	0	0	1	12	1	0	0	1	1	0	0	0
	Shanghai	4	2	21	9	4	9	10	20		750	313	1029	88	190	65	125	82	41	19	6	1	33	230	32	7	0	16	12	2	2	13
	Jiangsu	4	2	13	7	6	9	10	21	48		192	1121	52	89	85	165	96	50	13	11	2	28	300	110	36	0	31	13	6	2	15
	Zhejiang	2	1	10	5	4	6	7	13	21	138		782	69	841	39	199	245	198	12	31	1	97	569	301	49	0	30	10	2	2	5
	Anhui	1	0	3	1	1	2	3	4	11	54	32		9	14	9	22	12	6	3	1	1	2	16	8	7	0	3	1	1	0	2
	Fujian	2	1	5	3	2	5	4	7	3	28	72	171		671	16	61	149	94	24	22	2	87	545	138	12	0	16	3	1	2	2
	Jiangxi	1	0	3	1	0	1	1	1	4	11	38	25	29		5	7	28	38	18	5	1	2	14	16	2	0	2	1	0	0	1
	Shandong	4	4	36	13	29	36	106	221	3	77	63	94	20	14		99	44	10	6	3	1	8	49	12	32	0	19	9	5	1	14
	Henan	3	1	25	23	4	5	5	11	1	29	39	51	12	9	38		59	12	9	2	1	7	58	5	4	1	38	9	5	1	7
	Hubei	2	1	9	5	1	3	3	3	3	23	45	32	25	34	7	111		61	18	6	2	55	124	9	6	1	13	3	4	0	3
	Hunan	1	0	4	2	1	2	1	3	1	6	25	11	17	28	3	13	95		39	11	2	10	39	19	7	0	3	2	1	0	2
	Guangdong	11	5	35	20	14	32	31	50	13	149	138	359	264	1611	99	1005	1464	3329		2213	90	323	2844	591	66	1	235	48	4	5	13
	Guangxi	1	0	3	1	1	2	2	3	1	6	26	14	17	18	3	9	20	132	57		6	6	40	42	13	0	2	1	0	0	1
	Hainan	1	0	2	1	1	3	3	5	1	4	7	11	12	22	5	14	38	53	58	55		10	56	9	2	0	4	1	0	0	1
	Chongqing	1	0	5	2	1	1	1	2	1	6	15	6	9	5	3	8	24	10	8	3	1		232	33	13	1	4	1	1	0	5
	Sichuan	4	2	13	8	3	5	3	5	9	14	40	10	17	9	8	17	24	16	57	6	2	122		24	43	9	17	11	7	1	30
	Guizhou	1	0	4	1	0	1	1	1	1	6	20	5	11	7	4	7	11	47	10	21	0	34	195		15	0	2	0	0	0	0

Yunnan	1	0	7	4	1	3	4	4	2	16	62	15	29	26	7	22	44	94	24	24	1	90	476	195		0	7	2	0	0	1
Tibet	0	0	0	0	0	0	0	0	0	1	2	1	1	0	1	3	2	2	0	0	0	5	68	0	1		3	10	5	0	0
Shaanxi	2	1	15	19	11	4	2	3	1	18	34	19	8	5	14	83	33	8	7	1	0	5	69	2	1	1		38	5	7	7
Gansu	1	1	7	3	2	2	1	1	1	14	22	9	5	3	8	34	11	4	2	0	0	2	31	1	0	1	31		15	6	8
Qinghai	0	0	3	2	1	1	0	0	0	8	6	5	2	1	4	18	3	2	1	0	0	1	20	0	0	1	11	31		1	1
Ningxia	0	0	4	2	8	2	1	2	0	5	9	11	2	1	6	24	2	1	0	0	0	1	13	0	0	0	40	56	1		1
Xinjiang	1	1	13	6	2	3	2	4	8	63	29	73	8	4	36	287	43	20	3	1	0	31	426	3	2	0	70	220	16	32	

Note: There were no zero interprovincial flows in 2000. The zero cells are the result of rounding for the purpose of presentation.

Table 3.7 Full interprovincial migration flow matrix in 2010 (Unit: 1000 persons)

		Origin																													
		Beijing	Tianjin	Hebei	Shanxi	Inner Mongolia	Liaoning	Jilin	Heilongjiang	Shanghai	Jiangsu	Zhejiang	Anhui	Fujian	Jiangxi	Shandong	Henan	Hubei	Hunan	Guangdong	Guangxi	Hainan	Chongqing	Sichuan	Guizhou	Yunnan	Tibet	Shaanxi	Gansu	Qinghai	Ningxia	Xinjiang
Destination	Beijing		83	1559	269	236	248	214	403	18	195	126	430	96	141	598	980	335	152	71	31	8	86	325	39	28	1	165	143	11	18	36
	Tianjin	23		754	101	85	82	102	194	5	72	43	139	29	34	505	331	108	34	19	9	3	27	109	17	11	1	64	64	7	6	10
	Hebei	75	65		58	103	75	60	196	2	37	32	63	20	20	89	181	62	25	9	9	3	24	83	18	15	1	44	20	5	4	7
	Shanxi	7	6	138		79	13	11	16	1	35	27	46	22	12	47	197	48	18	6	4	2	21	79	12	13	0	56	11	2	1	2
	Inner Mongolia	7	7	184	157		73	64	143	1	33	23	37	15	15	68	104	35	16	5	3	1	15	87	10	7	0	142	142	5	43	3
	Liaoning	9	8	69	17	180		300	570	4	44	30	75	20	13	120	122	35	17	11	4	2	22	69	8	6	1	15	9	3	2	4
	Jilin	4	4	20	6	35	64		131	2	14	14	20	7	5	45	26	15	6	4	2	2	4	15	2	2	0	4	3	1	1	2
	Heilongjiang	8	5	27	7	53	58	121		3	19	12	23	8	7	62	25	17	7	5	2	2	5	13	3	3	0	5	4	1	1	2
	Shanghai	23	13	67	45	24	63	59	99		1504	451	2602	264	487	378	783	408	229	79	49	9	228	624	148	70	1	126	95	11	10	29
	Jiangsu	14	9	84	51	21	37	40	69	83		268	2575	139	265	411	1016	401	203	45	46	6	179	651	276	141	2	207	102	14	7	21
	Zhejiang	8	6	42	25	14	28	32	55	39	342		2285	164	1530	185	1224	899	750	47	139	6	593	1241	1499	411	1	168	62	8	6	15
	Anhui	4	3	21	9	5	7	7	9	21	114	61		29	40	34	97	48	32	13	8	2	16	44	33	29	1	16	6	3	1	3
	Fujian	4	3	17	10	5	13	12	20	5	41	69	250		949	43	278	340	235	51	70	9	410	822	472	98	1	56	18	3	4	5
	Jiangxi	3	2	15	12	9	7	5	5	8	26	47	51	45		19	41	53	68	36	20	7	14	35	34	11	1	10	9	2	2	3
	Shandong	13	13	142	54	79	79	186	408	6	138	62	130	37	31		285	78	32	16	9	3	24	86	24	44	1	48	41	13	7	24
	Henan	7	5	39	36	8	9	8	15	4	32	34	61	20	19	49		60	25	17	7	3	14	39	11	9	1	30	12	7	2	8
	Hubei	7	4	26	16	11	9	6	10	5	38	57	60	37	58	30	190		98	40	20	8	112	74	23	17	1	27	13	5	3	9
	Hunan	5	3	16	10	5	7	7	7	3	18	34	25	33	72	18	46	130		67	34	8	23	51	46	21	1	15	9	3	2	6
	Guangdong	19	12	86	55	25	66	74	109	14	126	156	445	434	1871	156	1762	2335	4602		3555	164	934	2602	958	322	1	439	125	13	9	26
	Guangxi	3	3	14	6	5	8	9	12	2	14	36	29	44	40	16	40	50	204	124		12	21	56	51	25	0	9	4	1	1	3
	Hainan	3	2	7	5	6	8	9	19	2	10	12	20	19	29	9	37	44	64	80	59		24	76	18	10	0	8	5	1	1	4

Chongqing	4	2	13	8	6	5	4	5	3	16	25	14	21	16	15	26	45	29	19	11	4		508	72	31	3	13	9	3	2	13
Sichuan	9	4	29	15	8	13	10	12	6	31	49	27	34	32	30	59	70	50	37	18	5	299		58	91	19	34	33	17	3	24
Guizhou	1	1	9	4	1	3	4	4	2	15	32	19	29	22	13	28	41	115	18	26	2	109	209		40	0	10	3	1	0	1
Yunnan	2	2	12	6	3	7	6	8	3	16	46	23	41	36	15	42	57	133	25	34	3	153	345	187		1	14	6	3	1	4
Tibet	0	0	1	1	0	0	0	0	0	1	2	3	1	1	1	9	5	4	0	0	0	11	84	1	4		5	20	8	0	1
Shaanxi	7	4	39	67	36	12	8	12	3	34	35	42	26	18	41	160	64	26	14	6	2	25	117	11	7	6		103	12	24	16
Gansu	3	3	14	8	6	6	4	4	2	19	23	15	10	7	15	57	26	12	6	2	1	9	44	3	3	3	48		58	11	12
Qinghai	1	1	11	5	2	2	2	2	0	13	10	13	6	4	11	44	18	8	2	1	0	8	44	3	2	5	23	73		2	2
Ningxia	1	1	12	7	15	3	2	5	0	9	10	18	4	3	15	49	9	5	2	1	0	7	22	2	1	0	55	105	2		2
Xinjiang	2	2	28	15	6	8	6	11	5	52	30	81	15	10	56	387	53	32	12	4	1	90	351	10	10	0	102	344	18	50	

Note: There were no zero interprovincial flows in 2010. The zero cells are the result of rounding for the purpose of presentation.

Chapter 4 Analysis of the rural and urban income divide and interprovincial migration in China from 2000 to 2010 with gravity models

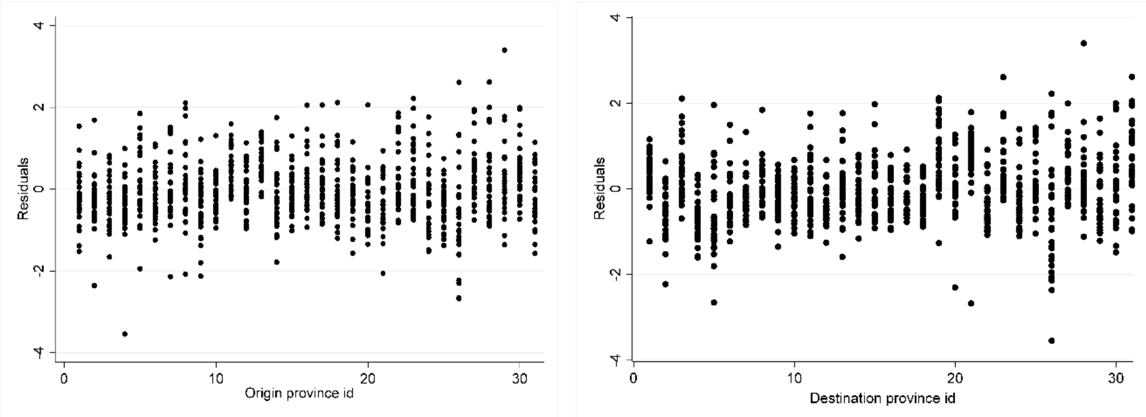

Figure 4.5 Clustering of urban-urban migration flow residuals against origin and destination provinces

Figure 4.5 shows how residuals cluster against both origins and destinations for urban-urban migration flows. Residuals of other three migration types have similar patterns so will not be presented here.

Chapter 5 Analysing interprovincial urban migration flows in China: A new multilevel gravity model approach

5.6 Derivations for the four model implied correlations

This Sub-section presents the derivations for the four model implied correlations (Type 1, 2, 3 and 4) presented in Table 5.1. Reconsider Model 3 (Equation 5.2), the most general model presented in Section 5.3. The conditional variance for migration flow y_{ij} is given by

$$\mathrm{Var}\left(y_{ij}\middle|\mathbf{x}_{1i},\mathbf{x}_{2j},\mathbf{x}_{3ij}\right)=\mathrm{Var}\left(o_i+d_j+e_{ij}\right)=\sigma_o^2+\sigma_d^2+\sigma_e^2 \tag{5.3}$$

while the conditional covariance between two migration flows y_{ij} and $y_{i'j'}$ is given by

$$\mathrm{Cov}\left(y_{ij},y_{i'j'}\middle|\mathbf{x}_{1i},\mathbf{x}_{2j},\mathbf{x}_{3ij},\mathbf{x}_{1i'},\mathbf{x}_{2j'},\mathbf{x}_{3i'j'}\right)=\mathrm{Cov}\left(o_i+d_j+e_{ij},o_{i'}+d_{j'}+e_{i'j'}\right) \tag{5.4}$$

and will vary in strength depending on the extent to which the origin and destination provinces of each flow overlap or not. The conditional correlation can then be calculated in the usual way as follows

$$\mathrm{Corr}\left(y_{ij},y_{i'j'}\middle|\mathbf{x}_{1i},\mathbf{x}_{2j},\mathbf{x}_{3ij},\mathbf{x}_{1i'},\mathbf{x}_{2j'},\mathbf{x}_{3i'j'}\right)=\frac{\mathrm{Cov}\left(o_i+d_j+e_{ij},o_{i'}+d_{j'}+e_{i'j'}\right)}{\sqrt{\mathrm{Var}\left(o_i+d_j+e_{ij}\right)}\sqrt{\mathrm{Var}\left(o_{i'}+d_{j'}+e_{i'j'}\right)}} \tag{5.5}$$

The four types of correlation presented in Table 5.1 can then be derived as follows:

Type 1: The correlation between the flow from origin i to destination j and from origin i to destination k is given by

$$\mathrm{Corr}\left(y_{ij},y_{ik}\middle|\mathbf{x}_{1i},\mathbf{x}_{2j},\mathbf{x}_{3ij},\mathbf{x}_{1i},\mathbf{x}_{2k},\mathbf{x}_{3ik}\right)=\frac{\sigma_o^2}{\sigma_o^2+\sigma_d^2+\sigma_e^2} \tag{5.6}$$

Type 2: The correlation between the flow from origin i to destination j and from origin j to destination k is given by

$$\mathrm{Corr}\left(y_{ij}, y_{jk} \middle| \mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3j}, \mathbf{x}_{1j}, \mathbf{x}_{2k}, \mathbf{x}_{3jk}\right) = \frac{\sigma_{od}}{\sigma_o^2+\sigma_d^2+\sigma_e^2} \tag{5.7}$$

Type 3: The correlation between the flow from origin i to destination j and from origin k to destination j is given by

$$\mathrm{Corr}\left(y_{ij}, y_{kj} \middle| \mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3ij}, \mathbf{x}_{1k}, \mathbf{x}_{2j}, \mathbf{x}_{3kj}\right) = \frac{\sigma_d^2}{\sigma_o^2+\sigma_d^2+\sigma_e^2} \tag{5.8}$$

Type 4: The correlation between the flow from origin i to destination j and from origin j to destination i is given by

$$\mathrm{Corr}\left(y_{ij}, y_{ji} \middle| \mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3ij}, \mathbf{x}_{1j}, \mathbf{x}_{2i}, \mathbf{x}_{3ji}\right) = \frac{2\sigma_{od}+\sigma_{ee}}{\sigma_o^2+\sigma_d^2+\sigma_e^2} \tag{5.9}$$

Model 1 (Equation 3.3) is a constrained version of Model 3 where $\sigma_o^2 = 0$, $\sigma_d^2 = 0$, $\sigma_{od} = 0$ and $\sigma_{ee} = 0$ and so all four correlations are implicitly assumed to be zero. Model 2 (Equation 5.2) is a constrained version of Model 3 where $\sigma_{od} = 0$ and $\sigma_{ee} = 0$ and so the last two correlations are implicitly assumed to be zero.

5.7 Results comparison of additional models

Table 5.4 presents all eight models used in Chapter 5. All five intermediate models can be classified into three groups: models of the first group have two levels and focus on the investigation of origin (Model 1a) or destination (Model 1b) effect only, whose main purpose is to compare with the base model of Model 1 to check whether clustering effects of origins or destination are important; in order to better understand Model 2, the second group of models are also constructed as two-level cross-classified by investigating the effects of origin and destination simultaneously, with Model 2a exploring equal origin and destination effects and Model 2b measuring correlation of origin and destination effects respectively; and Model 3a is the only member in the third model group to check the robustness of the proposed Model 3 by exploring whether origin and destination effects can be equal. By conducting such systematic comparisons of models, a better understanding of the measurement development process can be achieved, and details and nuances about effects of flow-pairs, origins and destinations upon interprovincial migration in China could be identified more clearly.

Table 5.4 All models used in Chapter 5

	Model 1	Model 1a	Model 1b	Model 2	Model 2a	Model 2b	Model 3	Model 3a
Number of levels	1	2	2	2	2	2	3	3
Variance constraints	σ_o^2=0 σ_d^2=0 σ_e^2≠0	σ_o^2≠0 σ_d^2=0 σ_e^2≠0	σ_o^2=0 σ_d^2≠0 σ_e^2≠0	σ_o^2≠0 σ_d^2≠0 σ_e^2≠0	σ_o^2≠0 σ_d^2≠0 σ_o^2=σ_d^2 σ_e^2≠0	σ_o^2≠0 σ_d^2≠0 σ_e^2≠0	σ_o^2≠0 σ_d^2≠0 σ_e^2≠0	σ_o^2≠0 σ_d^2≠0 σ_o^2=σ_d^2 σ_e^2≠0
Covariance constraints	σ_{od}=0 σ_{ee}=0	σ_{od}=0 σ_{ee}=0	σ_{od}=0 σ_{ee}=0	σ_{od}=0 σ_{ee}=0	σ_{od}≠0 σ_{ee}=0	σ_{od}≠0 σ_{ee}=0	σ_{od}≠0 σ_{ee}≠0	σ_{od}≠0 σ_{ee}≠0

Note: σ_e^2 is pair flow level variance; σ_o^2 and σ_d^2 are origin/destination variances respectively; σ_{od} is the origin and destination covariance, whilst σ_{ee} is the flow-pair covariance.

Table 5.5 presents results of all eight models. Coefficients and their standard errors see substantive changes across models. Moving from Model 1 to the first group of intermediate models, much improvement is observed in model fit (Model 1a is preferred to Model 1, likelihood ratio test $\chi_1^2 = 79.0$, $p < 0.001$; Model 1b is also preferred to Model 1, likelihood ratio test $\chi_1^2 = 240.6$, $p < 0.001$). Model 1a sees notable changes in coefficients and standard errors of the constant, and urban population and income in the origin, whilst Model 1b sees obvious changes in coefficients and standard errors of all covariates.

As mentioned in Section 5.4 of Chapter 5, Model 2 is significantly preferred to Model 1. However, moving from Model 2 to the second group of intermediate models does not see significant improvement in model fit (likelihood ratio test of Model 2 and 2a, $\chi_1^2 = 3.7$, $p > 0.05$; likelihood ratio test of Model 2 and 2b, $\chi_1^2 = 0.8$, $p > 0.05$). There are no observable changes in coefficients and standard errors of any covariates either. Although Model 3 is significantly preferred to Model 2 (Section 4.4 of Chapter 4), moving from Model 3 to Model 3a does not see either significant improvement in model fit (likelihood ratio test of Model 3 and 3a, $\chi_1^2 = 3.8$, $p > 0.05$) or any observable changes in coefficients and standard errors of covariates.

Table 5.5 Results comparison of all models

		Model 1	Model 1a	Model 1b	Model 2	Model 2a	Model 2b	Model 3	Model 3a
		Estimate(SE)	Estimate(SE)	Estimate(SE)	Estimate(SE)	Estimate(SE)	Estimate(SE)	Estimate(SE)	Estimate(SE)
Fixed part									
	constant(cons)	4.995(0.613)***	5.177(0.887)***	6.712(1.153)***	7.534(1.396)***	7.518(1.396)***	7.575(1.492)***	7.410(1.503)***	7.372(1.501)***
	origin urban population(lnoupop)	0.921(0.032)***	0.921(0.070)***	0.913(0.027)***	0.909(0.075)***	0.909(0.092)***	0.909(0.075)***	0.909(0.074)***	0.910(0.091)***
	destination urban population(lndupop)	0.632(0.032)***	0.631(0.030)***	0.623(0.102)***	0.619(0.106)***	0.619(0.092)***	0.619(0.106)***	0.620(0.105)***	0.620(0.091)***
	origin urban income(lnouincome)	-0.990(0.122)***	-0.995(0.268)***	-1.043(0.103)***	-1.068(0.284)***	-1.067(0.349)**	-1.069(0.284)***	-1.064(0.283)***	-1.063(0.347)**
	destination urban income(lnduincome)	2.182(0.122)***	2.177(0.114)***	2.129(0.387)***	2.104(0.403)***	2.105(0.349)***	2.103(0.404)***	2.108(0.401)***	2.109(0.347)***
	distance(lndistance)	-0.818(0.044)***	-0.838(0.044)***	-1.012(0.040)***	-1.105(0.039)***	-1.104(0.039)***	-1.110(0.039)***	-1.091(0.051)***	-1.087(0.051)***
Random part									
Level 3									
	origin province variance(σ_o^2)		0.085(0.026)		0.106(0.030)	0.165(0.032)	0.106(0.030)	0.104(0.030)	0.162(0.032)
	destination province variance(σ_d^2)			0.202(0.055)	0.224(0.060)	0.165(0.032)	0.225(0.060)	0.221(0.060)	0.162(0.032)
	origin-destination covariance(σ_{od})						0.026(0.030)	0.016(0.030)	0.015(0.032)
	origin-destination correlation(ρ_{od})						0.172(0.190)	0.105(0.193)	0.093(0.192)
Level 2									
	flow pair covariance(σ_{ee})							0.250(0.021)	0.250(0.021)
	flow pair correlation(ρ_{ee})							0.719(0.023)	0.719(0.023)
Level 1									
	individual flow variance(σ_e^2)	0.641 (0.030)	0.556(0.026)	0.453(0.021)	0.348(0.017)	0.348(0.017)	0.348(0.017)	0.348(0.021)	0.348(0.021)

Chapter 6 Exploring the non-linearity of distance decay in interprovincial urban-urban migration flows

6.8 Results of additional models

Model 2a incorporates only the region dummies in the equation in order to test how the provincial-region effects alone affect the migration flows. Its equation can be written as

$$\mathbf{x}'_{1i} = [\ln(p_{ui}), \ln(I_{ui}), r_{1i}\ r_{2i}, r_{3i}]$$

$$\mathbf{x}'_{2j} = \left[\ln(p_{uj})\ , \ln(I_{uj}), r_{1j}, r_{2j}, r_{3j}\right]$$

$$\mathbf{x}_{3ij} = \left[\ln(d_{ij}), (\ln(d_{ij}))^2\ , r_{ij}\right] \qquad (6.7).$$

Model 2b is a simplified version of Model 3 by excluding the interaction effects between the neighbouring-province and provincial-region covariates. Its equation can be written as

$$\mathbf{x}'_{1i} = [\ln(p_{ui}), \ln(I_{ui}), r_{1i}, r_{2i}, r_{3i}]$$

$$\mathbf{x}'_{2j} = \left[\ln(p_{uj})\ , \ln(I_{uj}), r_{1j}, r_{2j}, r_{3j}\right]$$

$$\mathbf{x}_{3ij} = \left[\ln(d_{ij}), (\ln(d_{ij}))^2\ , n_{ij}, r_{ij}\right] \qquad (6.8).$$

Model 3a is also a modified version of Model 3 by adopting the traditional three-region definition in place of the four-region definition. In other words, provinces of the Northeast region are either treated as the East or the West. Thus, there are only two dummy variables r_{2i} and r_{3i} for the three origin regions and another two dummy variables r_{2j} and r_{3j} for the three destination regions. The reference category remains as the East region, and the subscripts 2 and 3 denote the remaining Central and West regions respectively.

$$\mathbf{x}'_{1i} = [\ln(p_{ui}), \ln(I_{ui}),\ r_{2i}, r_{3i}]$$

$$\mathbf{x}'_{2j} = \left[\ln\left(p_{uj}\right), \ln\left(I_{uj}\right), r_{2j}, r_{3j}\right]$$

$$\mathbf{x}_{3ij} = \left[\ln\left(d_{ij}\right), (\ln\left(d_{ij}\right))^2, n_{ij}, r_{ij}, n_{ij}r_{2i}, n_{ij}r_{3i}, n_{ij}r_{2j}, n_{ij}r_{3j}\right] \tag{6.9}$$

Table 6.6 Results from Model 2a, 2b and 3a

	Model 2a		Model 2b		Model 3a	
Variables	Estimate	SE	Estimate	SE	Estimate	SE
Fixed part						
Constant	3.322	3.379	-2.623	3.274	1.002	3.321
Log of origin urban population	0.910***	0.088	0.886***	0.084	0.894***	0.084
Log of destination urban population	0.788***	0.104	0.764***	0.101	0.750***	0.101
Log of origin urban income	-1.175**	0.459	-1.093*	0.436	-1.024*	0.426
Log of destination urban income	1.276*	0.54	1.358*	0.524	1.466**	0.508
Log of distance	1.055	0.74	2.441**	0.722	1.226	0.747
Square of log of distance	-0.167**	0.055	-0.251***	0.053	-0.163**	0.055
Neighbours (reference: Non-neighbours)			0.664***	0.087	0.318	0.169
Within-region (reference: Between-region)	-0.185**	0.065	-0.248***	0.061	-0.242***	0.053
Origin region (reference: East)						
Northeast	0.132	0.296	0.097	0.281		
Central	-0.226	0.248	-0.223	0.236	-0.037	0.222
West	-0.012	0.232	-0.081	0.221	-0.143	0.216
Neighbouring province*Origin region (reference: Neighbours*Origin East)						
Neighbours*Northeast						
Neighbours*Central					0.066	0.130
Neighbours*West					0.492***	0.128
Destination region (reference: East)						
Northeast	-0.494	0.347	-0.529	0.338		
Central	-1.023***	0.291	-1.020***	0.283	-0.755**	0.265
West	-0.096	0.273	-0.164	0.265	-0.192	0.258
Neighbouring province*Destination region (reference: Neighbours* Destination East)						
Neighbours*Northeast						
Neighbours*Central					-0.271*	0.130
Neighbours*West					0.288*	0.128
Random part						
Origin province variance	0.112***	0.032	0.101***	0.029	0.106***	0.030

Destination province variance	0.159***	0.044	0.150***	0.041	0.156***	0.042
Individual flow variance	0.333***	0.020	0.302***	0.017	0.287***	0.017
Origin-destination correlation	0.180	0.189	0.117	0.193	0.158	0.189
Flow-pair correlation	0.706***	0.024	0.675***	0.026	0.665***	0.028
Deviance	1472.3		1417.7		1387.9	
Total residual	0.604		0.553		0.549	

Note: Response variable is the log migration flow (in 1000s). *** denotes p < 0.001, ** denotes p < 0.01, and * denotes p < 0.05.

6.9 Additional figures of three other regions

Non-neighbouring flows Neighbouring flows

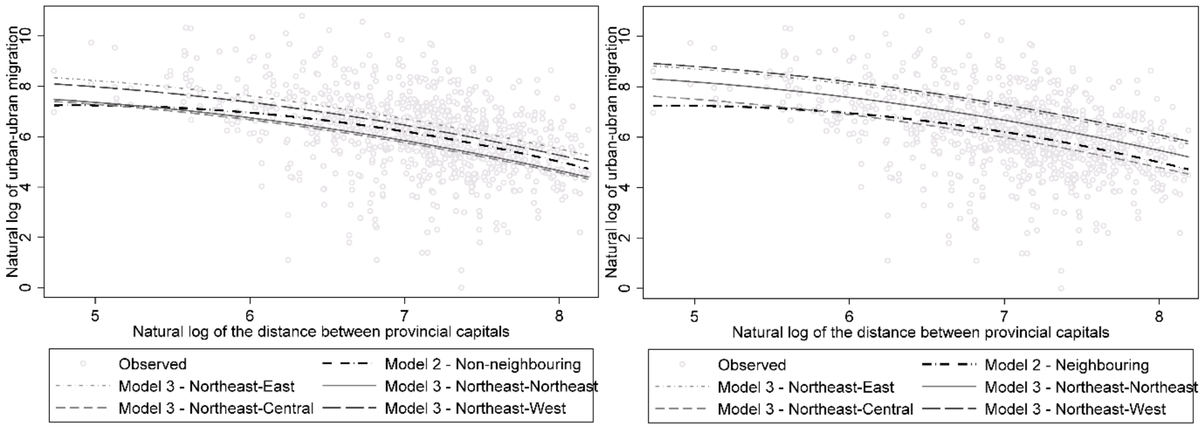

Figure 6.7 Comparison of fitted lines between Model 2 and 3 controlling the Northeast as the origin region

Non-neighbouring flows Neighbouring flows

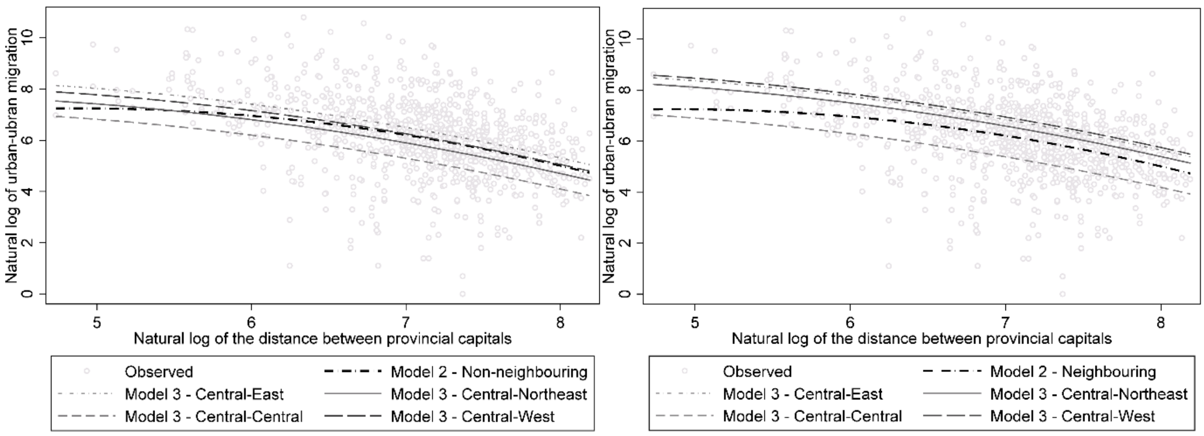

Figure 6.8 Comparison of fitted lines between Model 2 and 3 controlling the Central as the origin region

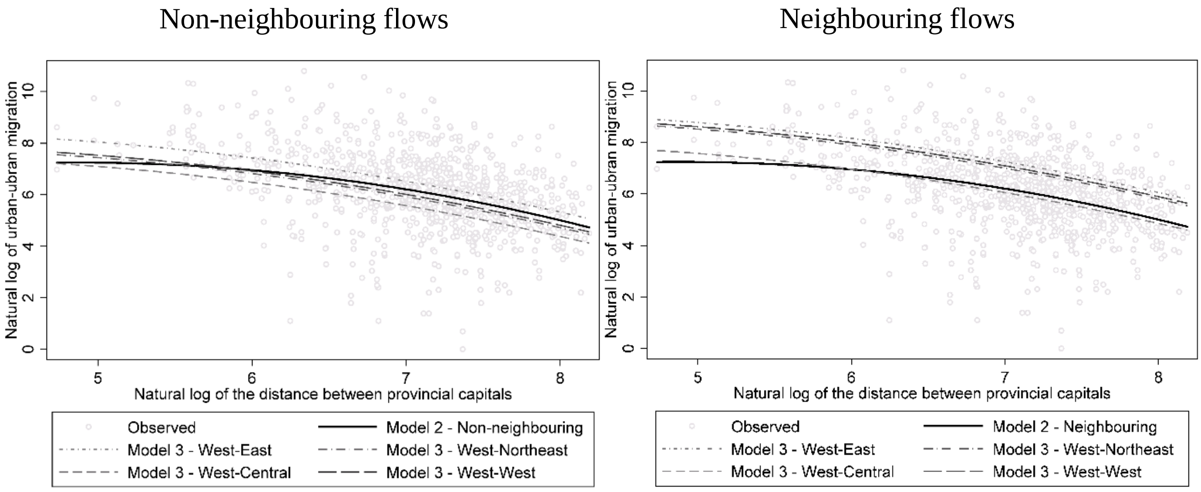

Figure 6.9 Comparison of fitted lines between Model 2 and 3 controlling the West as the origin region

6.10 Additional explanation of the random-part effects

The total residual variance ($\sigma_o^2 + \sigma_d^2 + \sigma_e^2$) also sees remarkable decreases moving from Model 1 to 3. Specifically, it falls from 0.678 in Model 1 to 0.636 in Model 2, and then stands at 0.528 in Model 3. This not only indicates the increase of overall model fit, but also the structural changes of VPCs. For instance, VPC of the destination province variance slightly increases from 33% in Model 1 to 35% in Model 2 before dropping to 27% in Model 3, whilst VPC of the individual flow variance dips from 50% in Model 1 to 49% in Model 2 before rising to 54% in Model 3. By contrast, VPCs of the origin province variance maintain within a relatively narrow range of 16% and 19%. To put it simply, much of the individual flow variance is mistakenly overestimated, by neglecting the neighbouring-province and provincial-region effects of flows.

Table 6.7 Flow dependencies of the three models

	Model 1		Model 2		Model 3	
Dependency (correlation conditional on the covariates)	Estimate	SE	Estimate	SE	Estimate	SE
Common origin (Type 1)	0.164***	0.041	0.160***	0.041	0.182***	0.045
Common destination (Type 2)	0.349***	0.063	0.347***	0.063	0.272***	0.057
Destination in first flow is origin in the second (Type 3)	0.046	0.046	0.023	0.046	0.026	0.044
Reciprocal flow (shared origin and destination) (Type 4)	0.440***	0.089	0.385***	0.093	0.420***	0.087

Note: Response variable is the log migration flow (in 1000s). *** denotes $p < 0.001$, ** denotes $p < 0.01$, and * denotes $p < 0.05$.

Table 6.7 presents the estimated correlations conditional on the covariates for the four forms of dependency, which experience similar patterns of change moving from Model 1 to 3 and further confirm findings of the random effects. Note that the definition of the residual flow has changed from Model 1 to 3 as new covariates are added, so the correlations between residual flows also change accordingly. To be more specific, the four forms of dependency become relevant to the region effects in Model 3. For instance, the model-implied correlation of flows sharing a common origin (Type 1) remain within the range 0.16~0.18 moving from Model 1 to 3, emphasising the relative stability of Type 1 dependency. This is because the estimated origin effect has remained relatively stable as mentioned earlier regardless of all the new covariates being added, meaning that the variation in province exporting power is not substantively captured by either the neighbouring-province relationship, the region effects or their interactions. By contrast, the correlation of flows sharing a common destination (Type 2) undergoes greater changes, by decreasing from 0.35 of Model 1 and 2 to 0.27 in Model 3. In other words, Type 2 dependency reduces from what is more than twice of that of Type 1 in Model 1 to the near equivalence in Model 3. This is related to the much-decreased destination effect from Model 1 to 3, indicating that a big component of the variation in province pulling power is captured by the regions wherein provinces are located along with the interactions with the neighbouring-province relationship. Meanwhile, the correlation between two residual flows where the destination of the first flow is the origin of the second (Type 3) also experiences some falling but remains insignificant across models. The overall falling pattern also holds true for the correlation between reciprocal residual flows (Type 4), which drops from 0.44 in Model 1 to 0.39 in Model 2 and then increases to 0.42 in Model 3. This shows that some of the flow-pair correlation is explained way by the region effect and its interactions with the neighbouring-province covariate. For instance, having adjusted for populations, incomes, the provincial-capital distance and the within-region effect, provinces in the West have larger both in- and out-flows between their neighbouring provinces than provinces in the east.

6.11 Additional province analysis

This sub-section will explain the origin and destination province effects first, and then present the correlation between them across all four models from Model 0 to Model 2a. In this way, it is possible to fully investigate the empirical meaning of the random-part effects on the provincial level.

6.11.1 Origin province effect

Figure 6.10 shows the residual differences between the origin provinces in the original measurement units for all the four models. By following the same procedure in Chapter 5, the provinces have been put into three groups based on whether the 95% confidence intervals overlap the reference line or not.

However, the interpretation of origin province effect in Model 3 becomes relative to the region average rather than the China wide average after introducing region covariates in the model. In other words, unlike in other models, origin province effects of Model 3 are divided into three groups, depending on whether their 95% confidence intervals overlap with the overall regional rather than national average or not. For instance, having adjusted for populations, incomes, distances and after controlling provincial neighbouring relationship along with region effects, Model 3 reveals three provinces (Zhejiang, Fujian and Ningxia) that are distinctively different from the rest of provinces. To be more specific, they depart significantly from what the theoretical model predicts by systematically exporting more urban-urban migrants than their regional average respectively at the 95% confidence level.

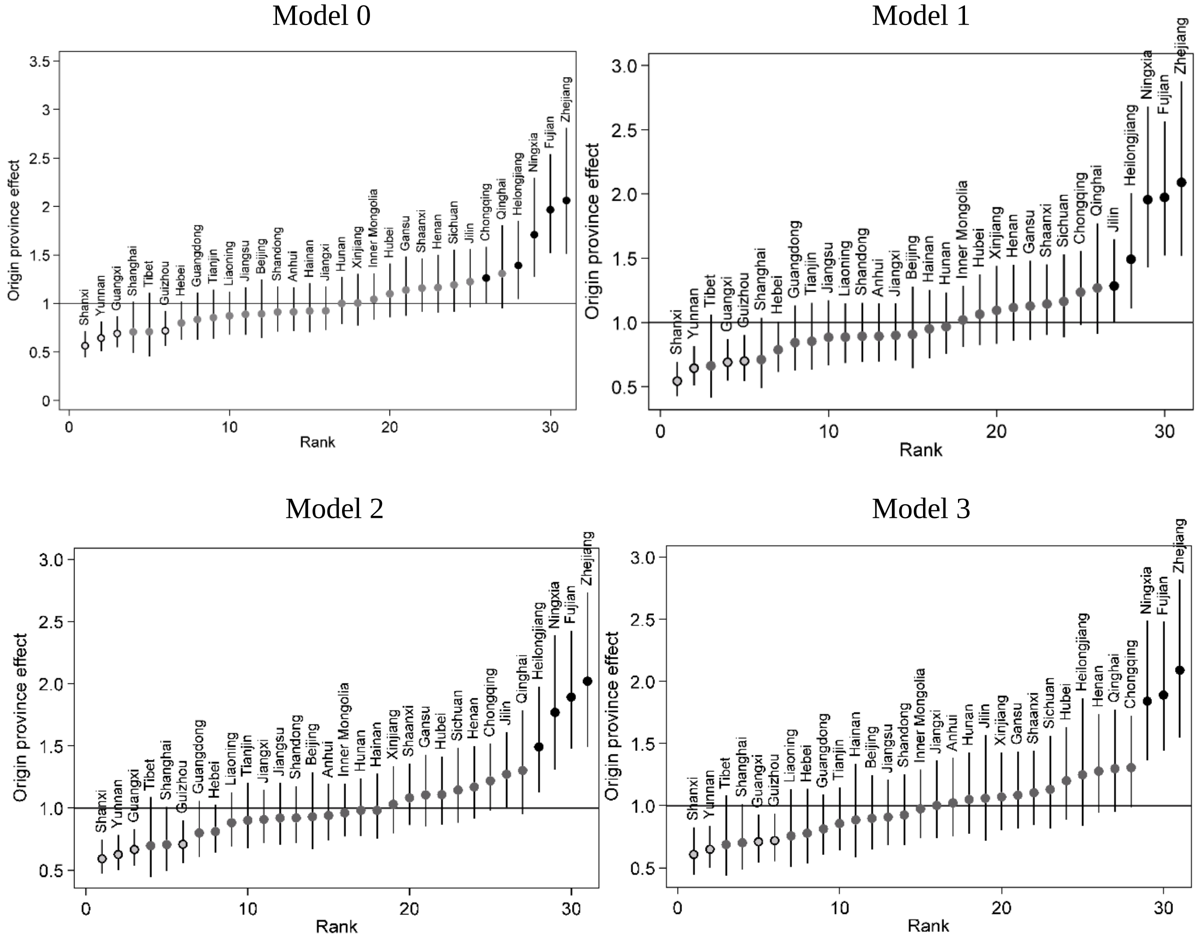

Figure 6.10 Predicted origin province effects plotted in rank order with 95% confidence intervals (unit: thousands)

In addition, all four models of this chapter are similar in both their values and rankings of the provinces regarding the origin province effect, whilst some variations can be observed (Figure 6.10). Specifically, Jilin changes from being not significantly different from the national average to significantly above the national average by introducing the quadratic log distance term in the model, moving from Model 0 to Model 1 of this Chapter. However, Jilin changes back to being not significantly different from the national average moving from Model 1 to 2. Moving from Model 2 to Model 3 sees no further changes other than Heilongjiang degrading from being significantly above the national average to being not significantly different from the regional average. This further confirms some component of the origin province effect for Heilongjiang is captured by the Northeast within which it is located, evidencing the need to consider the region effects in the model.

6.11.2 Destination province effect

The destination province effect can be considered in the same manner as the origin province effect as shown by Figure 6.11. Likewise, the interpretation of the destination province effect in Model 3 becomes relative to the region average rather than the China wide average.

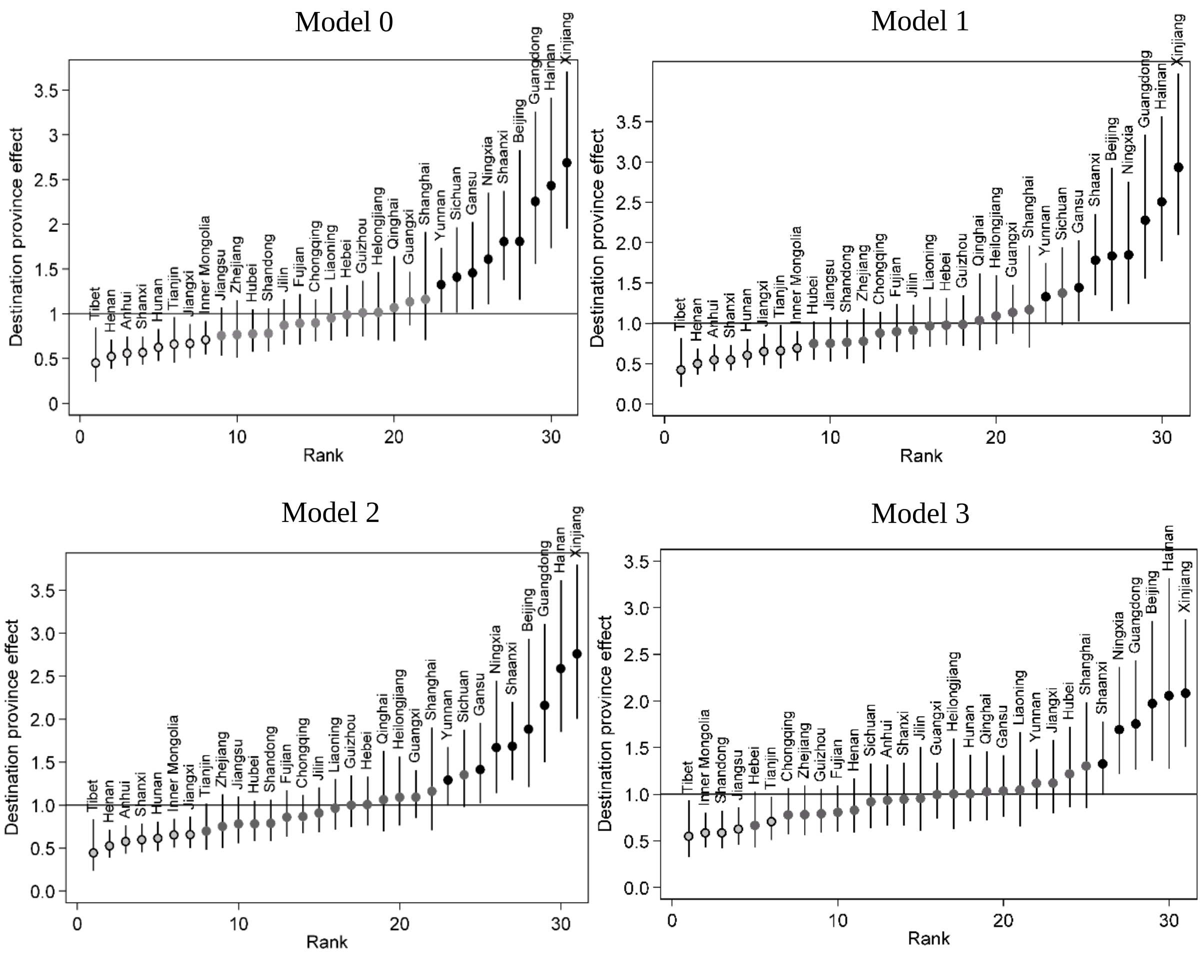

Figure 6.11 Predicted destination province effects plotted in rank order with 95% confidence intervals (unit: thousands)

As expected in Figure 6.11, the number of provinces with significant destination effects (including both above- and below-average) exceeds that of the origin, compared within each model respectively. For instance, having adjusted for populations, incomes, distances and after controlling provincial neighbouring relationship along with region effects, Model 3 has 11 and 7 provinces with significant destination and origin effects respectively. This re-confirms that the destination province effect is not only bigger than the origin province effect but also more variable than the latter in general.

Overall, all models are similar in both their values and rankings of the provinces regarding the destination province effect, but variations across models are observable. Specifically, Sichuan changes from being significantly above the national average to not significantly different from the national average by introducing the quadratic log distance term in the model, moving from Model 0 to Model 1. Moving from Model 1 to Model 2, Tianjin changes from being significantly below the national average to not significantly different from the national average by further introducing the neighbouring-province covariate in the model.

However, moving from Model 2 to Model 3 sees more changes. First, the number of provinces with significant below-average destination effects declines from 7 to 5, whilst that of above-average destination effects decreases from 8 to 6. Secondly, the rankings undergo radical changes. To begin with, 2 provinces (Gansu and Yunnan, both are in the West) change from being significantly above the national average to not significantly different from the regional average. Meanwhile, 5 provinces (Henan, Anhui, Shanxi, Hunan and Jiangxi, all are from the Central region) change from being significantly below the national average to not significantly different from the regional average. By contrast, 3 provinces (Shandong, Jiangsu and Tianjin, all are in the East) experience changes in the revert direction, by changing from not significantly different to being below the national average significantly from the regional average. Compared with the origin province effect, this re-confirms that some component of the destination province for many provinces is captured by the region within which they are located, which is also in line with the relatively bigger and more viable destination effect.

6.11.3 Correlation of origin and destination province effects

Table 6.8 presents the correlation of origin and destination province effects across Model 1 and 3. In general, the correlations between origin and destination province effects are small and insignificant both within and across all three models. By contrast, correlations between the origin province effects are large and significant ranging from 0.97 to 0.99 across three models, so are the correlations between the destination province effects but with a slightly wider range between 0.79 and 0.99.

Table 6.8 Correlation of origin and destination province effects

		Model 1		Model 2		Model 3	
		Origin	Destination	Origin	Destination	Origin	Destination
Model 1	Origin						
	Destination	0.021					
Model 2	Origin	0.993***	-0.016				
	Destination	0.140	0.794***	0.075			
Model 3	Origin	0.968***	-0.028	0.972***	0.021		
	Destination	0.093	0.989***	0.051	0.797***	0.046	

Note: Response variable is the log migration flow (in 1000s). *** denotes $p < 0.001$, ** denotes $p < 0.01$, and * denotes $p < 0.05$.

Interestingly, the destination province effect of Model 2 has relatively smaller correlations with both Model 1 and 3, which is in line with the earlier finding of the comparatively bigger and more viable destination effect. This indicates that introducing the neighbouring-province covariate alone is not sufficient to fully capture the destination province effect, although the origin province effect suffers to a less extent. Indeed, the pulling and exporting capabilities of provinces receive unneglectable impacts from the regions that they are located in.